Limitations of Variational Quantum Algorithms: A Quantum Optimal Transport Approach

Giacomo De Palma,[1,*] Milad Marvian,[2,†] Cambyse Rouzé,[3,‡] and Daniel Stilck França[4,5,§]

[1]*Department of Mathematics, University of Bologna, Bologna 40126, Italy*

[2]*Department of Electrical & Computer Engineering and Center for Quantum Information and Control, University of New Mexico, Albuquerque, New Mexico 87131, USA*

[3]*Zentrum Mathematik, Technische Universität München, Garching 85748, Germany*

[4]*QMATH, Department of Mathematical Sciences University of Copenhagen, Universitetsparken 5, Copenhagen 2100, Denmark*

[5]*ENS Lyon, Inria, F-69342, Lyon Cedex 07, France*

The impressive progress in quantum hardware of the last years has raised the interest of the quantum computing community in harvesting the computational power of such devices. However, in the absence of error correction, these devices can only reliably implement very shallow circuits or comparatively deeper circuits at the expense of a nontrivial density of errors. In this work, we obtain extremely tight limitation bounds for standard noisy intermediate-scale quantum proposals in both the noisy and noiseless regimes, with or without error-mitigation tools. The bounds limit the performance of both circuit model algorithms, such as the quantum approximate optimization algorithm, and also continuous-time algorithms, such as quantum annealing. In the noisy regime with local depolarizing noise p, we prove that at depths $L = \mathcal{O}(p^{-1})$ it is exponentially unlikely that the outcome of a noisy quantum circuit outperforms efficient classical algorithms for combinatorial optimization problems like max-cut. Although previous results already showed that classical algorithms outperform noisy quantum circuits at constant depth, these results only held for the expectation value of the output. Our results are based on newly developed quantum entropic and concentration inequalities, which constitute a homogeneous toolkit of theoretical methods from the quantum theory of optimal mass transport whose potential usefulness goes beyond the study of variational quantum algorithms.

I. INTRODUCTION

The last years have seen remarkable progress in both the size and quality of available quantum devices, reaching the point that even the best classical computers cannot easily simulate them [1–4]. In spite of these achievements, current devices lack error correction and, thus, are inherently noisy. Considering the significant overheads required to implement error correction [5,6], this has raised the

_____

*giacomo.depalma@unibo.it

†mmarvian@unm.edu

‡cambyse.rouze@tum.de

§daniel.stilck_franca@ens-lyon.fr

quantum computing community's interest in investigating whether such noisy quantum devices can nevertheless outperform classical computers in tasks of practical interest [7].

One class of algorithms that is considered suited for this task is variational quantum algorithms [8,9]. In most cases, these hybrid quantum classical algorithms work by optimizing the parameters of a shallow quantum circuit to minimize a cost function [8,9]. Prominent examples of such algorithms include the variational quantum eigensolver [10] and the quantum approximate optimization algorithm (QAOA) [11]. As variational algorithms only require the implementation of shallow circuits and simple measurements, it was expected that they could unlock the computational potential of near-term devices.

However, recent results have highlighted several obstacles to achieving a practical quantum advantage through variational quantum algorithms. For instance, some works have shown that optimizing the parameters of the circuit is computationally expensive in various settings [12–15].

Other works have shown that constant depth quantum circuits cannot outperform classical algorithms for certain combinatorial optimization problems [16–19]. Furthermore, it has been observed [15,20,21] that such variational quantum algorithms are less robust to noise than previously expected: already a small density of errors is sufficient to ensure that classical algorithms outperform the noisy device.

In this article, we further investigate the limitations of variational quantum algorithms. Our contributions are twofold. First, we obtain extremely tight limitation bounds for standard noisy intermediate-scale quantum (NISQ) proposals in both the noisy and noiseless regimes, with or without error-mitigation tools. Second, we provide a new homogeneous toolkit of theoretical methods whose potential usefulness goes beyond the present topic of variational quantum algorithms. Our methods originate from the emerging field of quantum optimal transport [22–31]. As we will see, optimal transport techniques have the combined advantages of simultaneously simplifying, unifying, and qualitatively refining previously known statements regarding fundamental properties of the output state of shallow and noisy circuits.

A. Limitations of noisy variational quantum algorithms

More precisely, we obtain two new complementary sets of results providing a better understanding of the limitations of variational quantum algorithms both at very shallow depths, when the effect of noise is negligible, and for a small density of errors. In Sec. III, we first derive new properties for the output probability of (potentially noisy) shallow quantum circuits initiated in state $|0\rangle^{\otimes n}$ and after measurement in the computational basis. These findings directly improve upon celebrated recent results on the limitation of certain variational quantum algorithms to solve the max-cut problem for certain classes of bipartite D-regular graphs. We prove that QAOA requires at least logarithmic in system size depth L to outperform efficient classical algorithms in some instances [16,18,19,32]:

$$L \geq \frac{1}{2\log(D+1)} \log \frac{n}{576}. \tag{1}$$

We note that our bound in Eq. (1) exponentially improves upon the dependence on the degree D of the graph previously found in Ref. [16]. For instance, for $D = 55$ (the minimum value for which Bravyi *et al.* [16] could prove that shallow quantum circuits cannot outperform the classical algorithm by Goemans and Williamson), our bound implies that the QAOA requires a depth larger than 1 as soon as $n \gtrsim 10^6$, whereas Bravyi *et al.* [16] found that $n \gtrsim 10^{54}$.

Next, Sec. IV is concerned with the concentration profile of the output measure of noisy circuits at any depth

$L = \Omega(1)$ for simple noise models, e.g., layers of circuits interspersed by layers of one-qubit depolarizing noise of parameter p. For instance, with realistic depolarizing probability $p = 0.1$ applied independently to each qubit, we are able to prove that the number of vertices for the graph has to be smaller than 10^9 in order for the noisy algorithm to outperform the best known classical algorithm (see Theorem VI.2). Moreover, we prove that at depths $L = \mathcal{O}(p^{-1})$ it is exponentially unlikely that the outcome of a noisy quantum circuit outperforms efficient classical algorithms for combinatorial optimization problems like max-cut. Although previous results already showed that noisy quantum circuits are outperformed by classical algorithms at constant depth [20], the results only held for the expectation value of the output. In contrast, our methods imply that the probability of observing a single string with better energy than that outputted by an efficient classical algorithm is exponentially small in the number of qubits. This is a significantly stronger statement, although at the cost of slightly worse constants [20].

In addition, in Sec. V, we show that certain error-mitigation protocols cannot reverse our conclusions unless we allow for an exponential number of samples in the number of qubits. First, in Sec. V A we show that virtual distillation or cooling protocols [33,34] only have an exponentially small success probability at constant depth. Furthermore, for mitigation procedures that have as their goal to estimate expectation values of observables, we show stringent limitations at $\mathcal{O}(\log(n))$ depth in Sec. V B. At this depth, any error-mitigation procedure that takes as input $m = \text{poly}(n)$ copies of the output of a noisy quantum circuit is exponentially unlikely to yield an estimate that deviates significantly from the estimate we would obtain by providing m copies of a trivial product state as input. Thus, the copies of the noisy quantum circuit do not provide significantly more insights than sampling from trivial product states. Our results strengthen recent results on limitations of error mitigation [35,36] both in terms of the required depth for them to apply and by providing concentration inequalities instead of results in expectation.

B. Quantum optimal transport toolkit

The second main contribution of the present article is the development of a new set of simple methods from quantum optimal transport whose potential use is likely to exceed the problem of finding tighter limitations on variational quantum algorithms. Our first main tool leading to the results of Sec. III is an optimal transport inequality introduced by Milman [37] in his study of the concentration and isoperimetric profile of probability measures on Riemannian manifolds with positive curvature (see also Refs. [38–40] for discussions on some related optimal transport inequalities). Adapted to the present setting of n-bit strings $\{0, 1\}^n$ endowed with the

Hamming distance $d_H(x, y) := \sum_{i=1}^{n} |x_i - y_i|$, Milman's so-called $(2, \infty)$-*Poincaré inequality* is a property of a probability measure μ on the set $\{0, 1\}^n$ that asks for the existence of a constant $C > 0$ such that, for any function $f : \{0, 1\}^n \to \mathbb{R}$,

$$\text{Var}_\mu(f) \leq Cn\|f\|_L^2, \tag{2}$$

where $\|f\|_L := \sup_{x \neq y} |f(x) - f(y)|/d_H(x, y)$ denotes the Lipschitz constant of f with respect to the Hamming distance. Besides its natural application to bounding the probability that the function f deviates from its mean by means of Chebyshev's inequality, namely,

$$\mathbb{P}_\mu(|f - \mathbb{E}_\mu[f]| \geq \sqrt{n}r) \leq \frac{C\|f\|_L^2}{r^2}, \tag{3}$$

the $(2, \infty)$-Poincaré inequality further implies by duality the following symmetric concentration inequality: for any two sets $S_1, S_2 \subset \{0, 1\}^n$ such that $\mu(S_1), \mu(S_2) \geq \mu_0 > 0$,

$$d_H(S_1, S_2) \leq 3\sqrt{\frac{Cn}{\mu_0}}. \tag{4}$$

For instance, we prove in Proposition III.2 below that in the case of a noiseless circuit, the output measure μ_{out} satisfies the $(2, \infty)$-Poincaré inequality with constant $C = B^2$, where B denotes the light cone of the circuit, i.e., the maximal amount of output qubits being influenced by the value of an arbitrary input qubit through the application of the circuit. In that case, the resulting symmetric concentration inequality (4) quadratically improves over that previously derived in Ref. [32, Corollary 43]:

$$d_H(S_1, S_2) \leq 4\frac{\sqrt{n}B^{1.5}}{\mu_0}.$$

Moreover, the $(2, \infty)$-Poincaré inequality turns out to be a very simple and versatile tool compared to the nontrivial proof of Corollary 43 of Ref. [32], which required the use of Chebyshev polynomials and approximate projections. Moreover, it can be very easily adapted to noisy shallow quantum circuits and continuous-time local Hamiltonian evolutions. In this latter setting, it unifies and refines the main results of Ref. [19].

The tools described in the previous paragraph are adapted to the study of quantum circuits of depth $L = \mathcal{O}(\log(n))$ and related short-time continuous-time evolutions. In contrast, our second set of fundamental results in Sec. IV concerns the concentration profile of the output measure of noisy circuits at any depth $L = \Omega(1)$ for simple noise models, e.g., layers of circuits interspersed by layers of one-qubit depolarizing noise of parameter p. In this case, we appeal to recently developed tools such as contraction coefficients for sandwiched Rényi divergences

[41,42] to prove that the probability under the output measure μ_{out} of the circuit that an arbitrary n-bit function $f : \{0, 1\}^n \to \mathbb{R}$ deviates from its mean by a constant fraction an of the total number of qubits satisfies the sub-Gaussian property

$$\mathbb{P}_{\mu_{\text{out}}}(|f - \langle f \rangle_{\mu_{\text{out}}}| \geq an) \leq Ke^{-ca^2n/\|f\|_L^2} \tag{5}$$

for some constants $K, c > 0$ and $a \geq a_0 \geq 0$. Interestingly, such strong concentration inequalities are known to be equivalent to a strengthening of the $(2, \infty)$-Poincaré inequality (2) known as the *transportation-cost* inequality [43,44]. The latter states that, for any measure ν that is absolutely continuous with respect to μ,

$$W_1(\nu, \mu)^2 \leq C'nD(\nu\|\mu), \tag{6}$$

where $D(\nu\|\mu)$ denotes the relative entropy of ν with respect to μ, whereas

$$W_1(\nu, \mu) := \sup_{\|f\|_L \leq 1} (\mathbb{E}_\nu[f] - \mathbb{E}_\mu[f]) \tag{7}$$

is the Wasserstein distance of order 1 between ν and μ, also called the Monge-Kantorovich distance or earth mover's distance.

In summary, our results clearly illustrate the potential of optimal transport methods such as the $(2, \infty)$-Poincaré inequality and the stronger transportation-cost inequality to study the performance of variational algorithms. We also believe that the discussed methods can have broad applications beyond that of understanding the computational power and limitations of near-term quantum devices.

Indeed, variations of inequalities like those in Eqs. (2) and (6) have recently found applications in different areas of quantum information theory. For instance, in Ref. [45] they were used to obtain exponential improvements for the sample complexity in quantum tomography. In Ref. [26] they were used to derive concentration bounds for commuting Gibbs states and show a strong version of the eigenstate thermalization hypothesis, a topic of intense research in physics. Thus, we believe that the new inequalities and techniques developed here could pave the way to extending such results to larger classes of states.

II. NOTATION AND DEFINITIONS

In this section, we introduce the main concepts discussed in the rest of the paper. We also refer the reader to Appendix A for a complete list of notation.

A. Basic notions

Given a set V of $|V| = n$ qudits, we denote by $\mathcal{H}_V = \bigotimes_{v \in V} \mathbb{C}^d$ the Hilbert space of n qudits and by $\mathcal{B}(\mathcal{H}_V)$ the algebra of linear operators on $\mathcal{H}_V$. We denote by $\mathcal{O}_V$

the self-adjoint linear operators on $\mathcal{H}_V$, whereas $\mathcal{O}_V^T \subset \mathcal{O}_V$ is the subspace of traceless self-adjoint linear operators. By $\mathcal{O}_V^+$ we denote the subset of positive semidefinite linear operators on $\mathcal{H}_V$, and $\mathcal{S}_V \subset \mathcal{O}_V^+$ denotes the set of quantum states. Similarly, we denote by $\mathcal{P}_V$ the set of probability measures on $[d]^V$. For any subset $A \subseteq V$, we use the standard notation $\mathcal{O}_A, \mathcal{S}_A, \ldots$ for the corresponding objects defined on subsystem A. Given a state $\rho \in \mathcal{S}_V$, we denote by ρ_A its marginal on subsystem A. For any region $A \subset V$, the identity on $\mathcal{O}_A$ is denoted by $\mathbb{I}_A$, or more simply $\mathbb{I}$. Given an observable O, we define $\langle O \rangle_\sigma = \mathrm{tr}[\sigma O]$. We denote the probability of measuring an eigenvalue of O greater than $a \in \mathbb{R}$ in state σ as $\mathbb{P}_\sigma(O \geq a)$. Given two probability measures μ, ν over a common measurable space, $\mu \ll \nu$ means that μ is absolutely continuous with respect to ν and $d\mu/d\nu$ denotes the corresponding Radon-Nikodym derivative.

B. Wasserstein distance

We make extensive use of notions of quantum optimal transport. The Lipschitz constant of the self-adjoint linear operator $H \in \mathcal{O}_V$ is defined as [29, Section V]

$$\|H\|_L := 2 \max_{v \in V} \min_{H_{v^c}} \|H - H_{v^c} \otimes \mathbb{I}_v\|_\infty, \qquad (8)$$

where the infimum above is taken over operators $H_{v^c} \in \mathcal{O}_{V \setminus \{v\}}$ that do not act on v. Lipschitz observables, that is, those $H \in \mathcal{O}_V$ such that $\|H\|_L = \mathcal{O}(1)$, capture extensive properties of a quantum system. They include (i) few-body and/or geometrically local observables; (ii) quasilocal observables; and even (iii) observables of the form $O = \sum_{i=1}^{\tilde{n}} O_i$, where $\|O_i\|_\infty \leq 1$ and $\mathrm{supp}(O_i) \cap \mathrm{supp}(O_j) = \emptyset$ for $i \neq j$. It is worth mentioning that the latter are considered in the fundamental problem in quantum statistical mechanics regarding the equivalence between the microcanonical and canonical ensembles [26,46–48].

The quantum W_1 distance proposed in Ref. [29] admits a dual formulation in terms of the above quantum generalization of the Lipschitz constant: the quantum W_1 distance between states $\rho, \omega \in \mathcal{S}_V$ is expressed as [29, Section V]

$$W_1(\rho, \omega) = \max\{\mathrm{tr}[(\rho - \omega)H] : \|H\|_L \leq 1\}.$$

Whereas the trace distance measures the global distinguishability of states, the Wasserstein distance measures distinguishability with respect to extensive, quasilocal observables.

C. Local quantum channels

In this work, we consider evolutions provided with a local description.

Definition II.1: A (noisy) quantum circuit $\mathcal{N}_V$ on n qudits of depth L is a product of L layers $\mathcal{N}_1, \ldots, \mathcal{N}_L$, where each layer $\mathcal{N}_\ell$ can be written as a tensor product of quantum channels $\mathcal{N}_{\ell,e}$ acting on a set $e \subset V$ of vertices:

$$\mathcal{N}_V = \prod_{\ell \in [L]} \bigotimes_{e \in E_\ell} \mathcal{N}_{\ell,e} \equiv \prod_{\ell \in [L]} \mathcal{N}_\ell \qquad (9)$$

for some sets E_ℓ of disjoint subsets of vertices. The circuit is called unitary (or noiseless) whenever each of the channels $\mathcal{N}_{\ell,e}$ is unitary. We call set $\{E_l\}_{l \in [L]}$ the architecture of circuit $\mathcal{N}_V$ and write $E = \bigcup_\ell E_\ell$.

A key concept associated with the notion of a local evolution is that of a light cone. In the case of a quantum circuit, the light cone of a vertex $v \in V$ is the smallest set of vertices $I_v^{\mathcal{N}_V} \equiv I_v \subseteq V$ such that, for any observable $O \in \mathcal{O}_V$ such that $\mathrm{tr}_v O = 0$, we have $\mathrm{tr}_{I_v}(\mathcal{N}_V(O)) = 0$. We then denote the light cone of the circuit by $I_{\mathcal{N}_V} := \max_{v \in V} |I_v|$. In Sec. III B, we extend this notion to the case of a continuous-time Hamiltonian evolution, where light cones are defined thanks to Lieb-Robinson bounds [49].

III. CONCENTRATION AT THE OUTPUT OF SHORT-TIME EVOLUTIONS

In this section, we obtain concentration inequalities for the outputs of short-time evolutions. Our main tool is an inequality between the variance and the Lipschitz constant of an observable. For any $O \in \mathcal{B}(\mathcal{H}_V)$ and $\omega \in \mathcal{S}_V$, the variance of O in state ω is defined as

$$\mathrm{Var}_\omega(O) := \mathrm{tr}[\omega |O - \mathrm{tr}[\omega O]\mathbb{I}|^2] = \langle |O - \langle O \rangle_\omega \mathbb{I}|^2 \rangle_\omega.$$

We denote the Kubo-Martin-Schwinger (KMS) inner product associated with state σ as $\langle A, B \rangle_\sigma := \mathrm{tr}[A^\dagger \sigma^{1/2} B \sigma^{1/2}]$, and its corresponding norm as $\|H\|_\sigma$. We have, for any $H \in \mathcal{O}_V$ [50, Eq. (20)],

$$\|H - \mathrm{tr}[\sigma H]\mathbb{I}\|_\sigma^2 \leq \mathrm{Var}_\sigma(H). \qquad (10)$$

With a slight abuse of notation, we use the same terminology for the analogous functionals for classical probability distributions.

In analogy with the classical literature [37], we say that a state σ satisfies a $(2, \infty)$-Poincaré inequality of constant $C > 0$ if, for any $O \in \mathcal{O}_V$,

$$\mathrm{Var}_\sigma(O) \leq C|V| \|O\|_L^2. \qquad (11)$$

For instance, tensor product states $\rho \equiv \bigotimes_{v \in V} \rho_v$ satisfy the $(2, \infty)$-Poincaré inequality with constant $C = 1$ (see Appendix F). The main motivations for introducing these inequalities are the following direct consequences of the $(2, \infty)$-Poincaré inequality. We leave their proofs to Appendix E.

Theorem III.1. *Assume that state σ satisfies a $(2, \infty)$-Poincaré inequality of constant $C > 0$. Then the following statements hold.*

(1) *Noncommutative transport-variance inequality: for any two states $\rho_1, \rho_2 \in \mathcal{S}_V$ with corresponding densities $X_j := \sigma^{-1/2} \rho_j \sigma^{-1/2}$,*

$$W_1(\rho_1, \rho_2) \leq \sqrt{C|V|}(\|X_1 - \mathbb{I}\|_\sigma + \|X_2 - \mathbb{I}\|_\sigma).$$

(2) *Measured transport-variance inequality: denote by $\mu_\sigma \in \mathcal{P}_V$ the probability measure induced by the measurement of σ in the computational basis. Then, for any probability measure $\nu \ll \mu_\sigma$,*

$$W_1(\nu, \mu_\sigma) \leq \sqrt{C|V| \operatorname{Var}_{\mu_\sigma}(d\nu/d\mu_\sigma)}.$$

Moreover, for any two sets $A, B \subset [d]^V$, their Hamming distance $d_H(A, B)$ satisfies the following symmetric concentration inequality:

$$d_H(A, B) \leq \sqrt{C|V|}(\mu_\sigma(A)^{-1/2} + \mu_\sigma(B)^{-1/2}). \tag{12}$$

(3) *Concentration of observables: for any observable $O \in \mathcal{O}_V$ and $r > 0$,*

$$\mathbb{P}_\sigma(|O - \langle O \rangle_\sigma \mathbb{I}| \geq r) \leq \frac{C|V|\|O\|_L^2}{r^2}. \tag{13}$$

Note that the Wasserstein distance and the Lipschitz constant are invariant under product unitaries. Thus, the same results hold for measuring the state in any product basis, not necessarily only the computational basis.

Although it might not be obvious from the outset, the inequalities in item (2) are known to imply no-go results for outputs of shallow quantum circuits [16,19,32]. Consider, for example, the output distribution μ we obtain when measuring the Greenberger-Horne-Zeilinger (GHZ) state in the computational basis, i.e., the all-zeros or all-ones string. If we take A to contain the all-zeros string and B to contain the all-ones string, we clearly have $\mu(A) = \mu(B) = 0.5$ and $d_H(A, B) = n$. Thus, the GHZ state does not satisfy a $(2, \infty)$-Poincaré inequality with $C = \mathcal{O}(1)$.

A. Poincaré inequalities at the output of noisy circuits

We now bound constant C in various settings. It turns out that noisy shallow circuits satisfy a $(2, \infty)$-Poincaré inequality.

Proposition III.1. *For any tensor product input state ρ, output $\mathcal{N}_V(\rho)$ satisfies a $(2, \infty)$-Poincaré inequality with constant*

$$C \leq 4\left(I_{\mathcal{N}_V}^2 + \frac{\max_\ell |E_\ell|}{|V|} \sum_{\ell=1}^L \max_{e \in E_\ell} I(e, L - \ell)^2\right),$$

where, given a set $e \in E_\ell$ and $m \in \mathbb{N}$, $I(e, L - \ell)$ denotes the set of all vertices in V in the light cone of set e for the circuit constituted of the last $L - \ell$ layers of $\mathcal{N}_V$.

The proof of this proposition is deferred to Appendix F. When $\mathcal{N}_V \equiv \mathcal{U}_V$ is noiseless, we get the following tightening of Proposition III.1.

Proposition III.2. *For any tensor product input state ρ, output $\mathcal{U}_V(\rho)$ satisfies a $(2, \infty)$-Poincaré inequality with constant*

$$C \leq 4I_{\mathcal{N}_V}^2. \tag{14}$$

Note that, for any circuit, the light cone can grow at most exponentially in L.

B. Poincaré inequality for continuous-time quantum processes

We now consider the continuous-time setting, and restrict ourselves to a system whose interactions are modeled by a graph $G = (V, E)$ whose vertices V correspond to a system of $|V| = n$ qudits, and denote by $D := \max_{v \in V}\{v'|(v, v') \in E\}$ the maximum number of nearest neighbors to a vertex. In the (noiseless) continuous-time setting, one replaces the notion of a circuit by that of a local time-dependent Hamiltonian evolution.

Definition III.1: A (noiseless) continuous-time local quantum process is a unitary evolution $\{\mathcal{U}_V(t)\}_{t \geq 0}$ generated by the time-dependent Hamiltonian

$$H(t) = \sum_{e \in E} \alpha_e(t) H_e, \tag{15}$$

where H_e is a time-independent self-adjoint operator that acts nontrivially only on the edge $e \in E$ with norm $\|H_e\|_\infty \leq 1/2$. We also assume that $b := \sup_{t,e} |\alpha_e(t)| < \infty$ independently of the size of the system. In what follows, for any subregion $A \subset V$, we also denote the Hamiltonian restricted to A by $H_A(t) := \sum_{e \subset A} \alpha_e(t) H_e$, and its corresponding unitary evolution by $\{\mathcal{U}_A(t)\}_{t \geq 0}$.

For continuous-time unitary evolutions, the concept of the light cone is formalized by the existence of a Lieb-Robinson bound [49]. Since their introduction, Lieb-Robinson bounds have been extensively studied in various levels of generality for unitary [51] as well as dissipative Markovian evolutions [52]. In what follows, we define a distance $\operatorname{dist} : E \times E \to \mathbb{R}_+$ on the edge set E that, for any two edges $e = (v_1, v_2)$ and $e' = (v_1', v_2')$, takes the value $\operatorname{dist}(e, e') = 0$ if and only if $e = e'$, and is otherwise equal to the length of the shortest path connecting the sets of vertices $\{v_1, v_2\}$ and $\{v_1', v_2'\}$. Next, we denote by $S_e(k)$ the sphere around any edge $e \in E$ of radius k, i.e.,

$$S_e(k) := \{e' \in E : \operatorname{dist}(e, e') = k\}. \tag{16}$$

Then, set E is said to be of spatial dimension δ if there is a constant $M > 0$ such that, for all $e \in E$, $|S_e(k)| \leq M k^{\delta-1}$.

The following result is taken from Ref. [53, Theorem 2] (see also Ref. [19, Theorem 1] for a similar result).

Theorem III.2 (**Lieb-Robinson bound**). *Using the notation introduced above, for any subregions $A \subset B \subset V$ with $k_0 := \text{dist}(A, V\backslash B) \geq 2\delta - 1$, any state ρ, and $0 \leq t$,*

$$\|\text{tr}_{A^c}(\mathcal{U}_V(t)(\rho) - \mathcal{U}_B(t)(\rho))\|_1 \leq \frac{2M}{2D-1} k_0^{\delta-1} e^{vt-k_0},$$

where $v := eb(2D-1)$ is the Lieb-Robinson velocity.

Next, we order the vertices $\{1, \ldots, n\}$, $n = |V|$, with their graph distance to an arbitrarily chosen vertex $v_0 \equiv 1$, and denote the graph distance $\text{dist}(\{1\}, \{i, \ldots, n\}) \equiv d(i)$. Then, using the notation of Sec. II, we have, for any $H \in \mathcal{O}_V$ (see Appendix G),

$$\|\mathcal{U}_V(t)^\dagger(H)\|_L$$

$$\leq \left(2(i_0 - 1) + \frac{4M}{2D-1} \sum_{i=i_0}^{|V|} d(i)^{\delta-1} e^{vt-d(i)}\right) \|H\|_L, \tag{17}$$

where i_0 stands for the first vertex such that $d(i_0) \geq 2\delta - 1$. By a reasoning that is identical to that leading to Proposition III.1, we have the following result.

Proposition III.3. *Let ρ be a product input state. For any $t \geq 0$, the output state $\mathcal{U}_V(t)(\rho)$ satisfies a $(2, \infty)$-Poincaré inequality with constant*

$$C_t \leq 4\left(2(i_0 - 1) + \frac{4M}{2D-1} \sum_{i=i_0}^{n} d(i)^{\delta-1} e^{vt-d(i)}\right)^2.$$

For a simpler version of the bound found in Proposition III.3, we refer the reader to Proposition VI.1 below. The bounds obtained in Proposition III.1, Proposition III.2, and Proposition III.3 can be combined with Theorem III.1(3) to get Chebyshev-type concentration bounds. This improves, for instance, over Theorem 2 of Ref. [19], where the concentration bound was obtained only in the continuous-time Hamiltonian setting and for a specific one-local observable measuring the Hamming weight. Moreover, the bound obtained in Eq. (12) on the Hamming distance between two sets in terms of their probabilities in state σ is an improvement over the symmetric concentration inequality found in Ref. [32, Corollary 43], namely,

$$d_H(A, B) \leq 4|V|^{1/2} I_{\mathcal{U}_V}^{3/2} \max\{\mu_\sigma(A)^{-1}, \mu_\sigma(B)^{-1}\}, \tag{18}$$

as well as its continuous-time analogue in Ref. [19, Theorem 3]. In summary, the $(2, \infty)$-Poincaré inequality is a versatile tool that we use to derive the strongest concentration-type bounds for general short-time quantum evolutions currently available in a simple, basis-free manner.

IV. LIMITATIONS AND CONCENTRATION INEQUALITIES FROM NOISE

In Sec. III we discussed how to use optimal transport methods to analyze the concentration profile of quantum circuits at small depths, even in the absence of noise. We now turn our attention to the case where the circuit is also subject to local noise and prove concentration inequalities for their outputs. As in the noiseless case, these can then be used to estimate the potential of noisy quantum circuits to outperform classical algorithms. However, unlike in Theorem III.1, we here obtain stronger Gaussian concentration inequalities.

For this, we make use of the sandwiched Rényi divergences [41,42] of order $\alpha \in (1, +\infty)$. For two states ρ, σ such that the support of ρ is included in the support of σ, they are defined as

$$D_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1} \log \text{tr}[(\sigma^{(1-\alpha)/2\alpha} \rho \sigma^{(1-\alpha)/2\alpha})^\alpha].$$

We also consider the relative entropy we obtain by taking the limit $\alpha \to \infty$,

$$D_\infty(\rho\|\sigma) = \log(\|\sigma^{-1/2} \rho \sigma^{-1/2}\|_\infty).$$

In case the support of ρ is not contained in that of σ, all the divergences above are defined to be $+\infty$.

We start from the assumption that the noise is driving the system to a quantum state σ on $\mathcal{H}_V$ that satisfies a Gaussian concentration inequality of parameter $c > 0$. That is, there is a constant K such that, for any $a > 0$ and observable O,

$$\mathbb{P}_\sigma(|O - \langle O \rangle_\sigma \mathbb{I}| \geq a|V|) \leq K e^{-ca^2|V|/\|\sigma^{-1/2}O\sigma^{1/2}\|_L^2}, \tag{19}$$

where the quantum Lipschitz constant of a non-self-adjoint matrix Z is defined as $\|Z\|_L := \max\{\|\text{Re}(Z)\|_L, \|\text{Im}(Z)\|_L\}$. Note that inequalities of the form (19) hold for product states [29,31,54], commuting high-temperature Gibbs states [26,55], and in slightly weaker form for all high-temperature Gibbs states [56] and gapped ground states on regular lattices [57]. Moreover, in the case where σ and O commute, we clearly have $\|\sigma^{-1/2}O\sigma^{1/2}\|_L = \|O\|_L$.

We then have the following concentration result, proved in Lemma B.1 in Appendix B.

Theorem IV.1. *Let σ satisfy Eq. (19). Then, for any state ρ and $a > 0$ and $\alpha > 0$, we have*

$$\mathbb{P}_\rho(|O - \langle O \rangle_\sigma \mathbb{I}| \geq a|V|)$$

$$\leq \exp\left(\frac{\alpha-1}{\alpha}\left(D_\alpha(\rho\|\sigma) - \frac{ca^2|V|}{\|\sigma^{-1/2}O\sigma^{1/2}\|_L^2} + \log(K)\right)\right). \tag{20}$$

It immediately follows that if we have, for a noisy circuit and a value of a,

$$\frac{D_\alpha(\mathcal{N}_V(\rho)\|\sigma)}{|V|} < \frac{ca^2}{\|\sigma^{-1/2}O\sigma^{1/2}\|_L^2} - \frac{\log(K)}{|V|}, \quad (21)$$

then the probability of observing an outcome outside of the interval $\langle O\rangle_\sigma \pm a|V|$ when measuring $\mathcal{N}_V(\rho)$ is exponentially small in $|V|$. Thus, given a bound on $D_\alpha(\mathcal{N}_V(\rho)\|\sigma)$, we can solve for a in Eq. (21) and establish a such that the probability of observing outcomes outside of $\langle O\rangle_\sigma \pm a|V|$ is exponentially small. In Sec. VI we discuss this more concretely to analyze the potential performance of QAOA under noise.

For now, let us discuss how to obtain the bounds on $D_\alpha(\mathcal{N}_V(\rho)\|\sigma)$ to effectively apply Theorem IV.1. One straightforward way to derive such bounds is to resort to so-called strong data-processing inequalities (SDPIs) [54,55,58–64]. A quantum channel $\mathcal{N}$ with fixed point σ is said to satisfy a SDPI with constant $q_\alpha > 0$ with respect to a fixed point σ and D_α if, for all other states ρ, we have

$$D_\alpha(\mathcal{N}(\rho)\|\sigma) \leq (1 - q_\alpha)D_\alpha(\rho\|\sigma). \quad (22)$$

Then, assuming that the noisy quantum circuit $\mathcal{N}_V$ we wish to implement is of the form (9) and each layer $\mathcal{N}_\ell$ satisfies Eq. (22) for some constant q_α, we show in Lemma C.1 in Appendix C that

$$D_\alpha(\mathcal{N}_V(\rho)\|\sigma) \leq (1 - q_\alpha)^L D_\alpha(\rho\|\sigma)$$
$$+ \sum_{\ell=0}^{L}(1 - q_\alpha)^{L-\ell}D_\infty\left(\bigotimes_{e\in E_\ell}\mathcal{N}_{\ell,e}(\sigma)\Big\|\sigma\right). \quad (23)$$

Thus, as long as the fixed point of the noise is left approximately invariant by the channels at the end of the circuit, Eq. (23) implies that the relative entropy will decay as the depth increases. As we argue in Appendix C 2, this will be the case for both QAOA and annealing circuits for most one-qubit noise models. Furthermore, this will hold for any circuit whenever the fixed point of the noise is the maximally mixed state.

It is also possible to derive similar inequalities for continuous-time evolutions with a time-dependent Hamiltonian H_t and the noise given by some Lindbladian $\mathcal{L}$. In that case, the assumption in Eq. (22) is replaced by

$$D_\alpha(e^{t\mathcal{L}}(\rho)\|\sigma) \leq e^{-r_\alpha t}D_\alpha(\rho\|\sigma) \quad (24)$$

for some constant $r_\alpha > 0$. In Lemma C.2 in Appendix C we show the continuous-time version of Eq. (23).

To illustrate the power of the bound in Eq. (21), let us analyze the case where $\mathcal{N}_V$ consists of a concatenation of layers of unitary gates with layers of noise $\mathcal{N} = \bigotimes_{k=1}^{|V|}\mathcal{D}_p$,

where $\mathcal{D}_p$ is a qubit depolarizing channel with depolarizing probability p. One can then show that Eq. (22) holds for $\alpha = 2$ and $q_2 = 2p + p^2$ [63, Sec. 3.3] and, thus, for any circuit of depth L in this noise model,

$$D_2\left(\mathcal{N}_V(\rho)\Big\|\frac{\mathbb{I}}{2^{|V|}}\right) \leq (1-p)^{2L}D_2\left(\rho\Big\|\frac{\mathbb{I}}{2^{|V|}}\right)$$
$$\leq (1-p)^{2L}|V|. \quad (25)$$

Moreover, the maximally mixed state satisfies Eq. (19) with $c = K = 1$ [30]. By combining Eq. (25) with Eq. (21) we arrive at the following result.

Proposition IV.1. *Let H be a traceless $|V|$-qubit Hamiltonian, and let $\mathcal{N}_V$ be a depth-L unitary circuit interspersed by one-qubit depolarizing noise with depolarizing probability p. Then, for any initial state ρ and $\epsilon > 0$,*

$$\mathbb{P}_{\mathcal{N}_V(\rho)}(|H| \geq ((1-p)^{2L} + \epsilon)^{1/2}\|H\|_L|V|)$$
$$\leq \exp\left(-\frac{\epsilon|V|}{2}\right). \quad (26)$$

Let us exemplify the power of Eq. (26). For an H of practical interest, say H is an Ising Hamiltonian, efficient classical algorithms are known to find solutions whose energy is a constant fraction from the ground-state energy [65]. That is, there exists an $a_c = \Omega(1)$ such that efficient classical algorithms can sample states ρ that satisfy $\text{tr}(\rho H) \leq -a_c|V|\|H\|_L$.

It then follows from Eq. (21) that at a constant depth $L > \log(a_c^{-1})/(2p)$, the probability of the noisy quantum circuit outperforming the classical algorithm is exponentially small in system size.

Note that other results in the literature already showed that quantum advantage is already lost at constant depth for such problems [20]. However, these results only showed bounds for the expectation value of the output of the circuit, whereas bounds like that in Proposition IV.1 provide concentration inequalities, a significantly stronger result. However, we do pay the price of having slightly worse constants for the depth at which advantage is lost compared to the results of Ref. [20]. We discuss concrete examples for the bounds we obtain on the depth in Sec. VI.

Above we illustrated our concentration bounds for depolarizing noise only, as it corresponds to the simplest noise model that we can analyze. But our result can be generalized to all noise models that contract the relative entropy uniformly with respect to a fixed point of full rank. However, this generalization comes at the expense of the bounds not being circuit independent unless the noise is unital. As before, the first step to obtain concentration results is to control the decay of the relative entropy under the noise for Rényi divergences.

Lemma IV.1 (Lemma 1 of Ref. [20]). *Let $\mathcal{N} : \mathcal{B}(\mathcal{H}_V) \to \mathcal{B}(\mathcal{H}_V)$ be a quantum channel with unique fixed point $\sigma > 0$ that satisfies a strong data-processing inequality with constant $p_\alpha > 0$ for some $\alpha > 1$. That is,*

$$D_\alpha(\mathcal{N}(\rho)\|\sigma) \le (1 - p_\alpha)D_\alpha(\rho\|\sigma) \qquad (27)$$

for all states ρ. Then, for any other quantum channels $\Phi_1, \ldots, \Phi_m : \mathcal{B}(\mathcal{H}_V) \to \mathcal{B}(\mathcal{H}_V)$, we have

$$D_\alpha\left(\prod_{t=1}^{m}(\Phi_t \circ \mathcal{N})(\rho)\Big\|\sigma\right)$$

$$\le (1 - p_\alpha)^m D_\alpha(\rho\|\sigma)$$

$$+ \sum_{t=1}^{m}(1 - p_\alpha)^{m-t}D_\infty(\Phi_t(\sigma)\|\sigma). \qquad (28)$$

We refer the reader to Appendix C for a more detailed discussion of this result and Lemma C.1 for a proof. In Appendix C 2 we evaluate the expression in Eq. (28) for the special case of QAOA circuits converging to diagonal product states. Furthermore, in Appendix C 3 we discuss the performance of the resulting bounds for random graphs.

In the same appendix we also prove the continuous-time version of the lemma above that is relevant to quantum annealers, which we now also state for completeness.

Proposition IV.2. *Let $\mathcal{L} : \mathcal{B}(\mathcal{H}_V) \to \mathcal{B}(\mathcal{H}_V)$ be a Lindbladian with fixed point σ_q defined as before with $q \ge 1/2$. Suppose that, for some $\alpha > 1$, we have, for all $t > 0$ and initial states, an $r_\alpha > 0$ such that*

$$D_\alpha(e^{t\mathcal{L}}(\rho)\|\sigma) \le e^{-r_\alpha t}D_\alpha(\rho\|\sigma). \qquad (29)$$

Moreover, for functions $f, g : [0,1] \to \mathbb{R}$ and $T > 0$, let $\mathcal{H}_t : \mathcal{B}(\mathcal{H}_V) \to \mathcal{B}(\mathcal{H}_V)$ be given by $\mathcal{H}_t(X) = i[X, f(t/T)H_X + g(t/T)H_I]$. Let $\mathcal{T}_t$ be the evolution of the system under the Lindbladian $\mathcal{S}_t = \mathcal{L} + \mathcal{H}_t$ from time 0 to $t \le T$. Then, for all states ρ,

$$D_\alpha(\mathcal{T}_T(\rho)\|\sigma) \le e^{-r_\alpha T}D_\alpha(\rho\|\sigma)$$

$$+ 2ne^{-r_\alpha T}\left(\sqrt{\frac{q}{1-q}} - \sqrt{\frac{1-q}{q}}\right)$$

$$\int_0^T e^{r_\alpha t}|f(t/T)|\,dt. \qquad (30)$$

Note that the expression in Eq. (28) will converge to 0 as long as $\Phi_t(\sigma) \simeq \sigma$ for t close to T. As we argue in more detail in Appendix C, this is expected to be satisfied for good QAOA circuits. Furthermore, we explicitly evaluate the bound in Eq. (28) in terms of the parameters of

the QAOA circuit in Corollary C.1 in Appendix C or for a given annealing schedule. These results can then be combined with Theorem IV.1 to understand the concentration properties of the output. The same holds in principle for Eq. (30), where this can be visualized more easily: as long as function f satisfies $f(1) = 0$, the second term in Eq. (30) will converge to 0.

We illustrate this concretely in the case of noisy annealers with a linear schedule in Proposition VI.3.

V. LIMITATIONS OF ERROR-MITIGATED NOISY VQAS: CONCENTRATION BOUNDS

A possible criticism of bounds like that of Proposition IV.1 is that they do not take error-mitigation techniques [66–69] into account. Although there does not seem to be a widely accepted definition of what error-mitigation entails, the overarching goal of such protocols is to extract information about noiseless circuits by sampling from noisy ones. Such proposals are expected to be useful before the advent of fault tolerance to reduce the level of noise present in the data outputted by NISQ devices.

The majority of existing mitigation protocols require a significant overhead in the number of samples to extract the noiseless signal from noisy ones, potentially making error mitigation prohibitively expensive. Thus, one of the main questions regarding the viability of error-mitigation strategies is the scaling of the sampling overhead they require in terms of the number of qubits, depth, and error rate.

There already exist some results in the literature discussing limitations of error mitigation, such as Refs. [35, 36]. They showed that certain error-mitigation protocols require a sampling overhead that is exponential in system size at linear circuit depth. Our results in the next sections suggest that at significantly lower depths it is already difficult to extract information about the noiseless output state, while also providing concentration bounds for error-mitigated circuits.

In what follows, we distinguish *sampling* and *weak* error mitigation. To the best of our knowledge, this distinction has not been made before in the literature. But in analogy with the terminology for the simulation of quantum circuits, we call an error-mitigation strategy a sampling protocol if it allows us to approximately sample from the output of a noiseless circuit. In contrast, weak error-mitigation techniques only allow for approximating expectation values of the outputs of noiseless circuits. Note that the latter is a weaker condition.

A. Sampling error mitigation and the effect of error mitigation on classical optimization problems

We start by discussing the effect of noise on known sampling error-mitigation procedures. We believe that these

are particularly relevant for classical combinatorial optimization problems. This is because, for such problems, one is often not necessarily interested in estimating the ground-state energy, but rather in obtaining a string of low energy that corresponds to a good solution. And for this, it would be necessary to obtain a sample.

To the best of our knowledge, the only error-mitigation technique that allows for sampling from the noiseless state is virtual distillation or cooling [33,34]. Going into the details of this procedure is beyond the scope of this manuscript. It suffices to say that it takes as an input k copies of the output of a noisy quantum circuit and aims at preparing state $\rho^k/\mathrm{tr}[\rho^k]$. Under some assumptions, one can then show that this state has an exponentially in k larger overlap with the output of the noiseless circuit. However, as this is clearly not a linear transformation, it can only be implemented stochastically. The success probability of the transformation is given $\mathrm{tr}[\rho^k] \leq \mathrm{tr}[\rho^2]$. As before, for simplicity, we state our no-go results for the case of local depolarizing noise and leave the proof and the more general case to Appendix C 2. We then have the following result.

Proposition V.1. *Let $\mathcal{N}_V$ be a depth-L unitary circuit interspersed by one-qubit depolarizing noise with depolarizing probability p. Then, for any initial state ρ and $k \geq 2$, the probability that virtual cooling or distillation succeeds is bounded by*

$$\mathrm{tr}[\mathcal{N}_V(\rho)^k] \leq \mathrm{tr}[\mathcal{N}_V(\rho)^2]$$
$$\leq \exp(-\log(2)(1-(1-p)^{2L})n). \quad (31)$$

The proof of Proposition V.1 can be found in Proposition D.1 in Appendix D. Thus, we conclude from Eq. (31) that unless the local noise rate is $p = \mathcal{O}(n^{-1})$, virtual distillation protocols will require an exponential in system size number of samples to be successful even after one layer of the circuit. We remark that our results essentially imply the same conclusions for general local, unital noise. For nonunital noise driving the system to a product state, we obtain the following statement.

Lemma V.1. *Let $\tau_q = q|0\rangle\langle 0| + (1-q)|1\rangle\langle 1|$, and assume without loss of generality that $q \leq 1/2$. Then, for any state $\rho \in \mathcal{S}_V$ with $n = |V|$ such that*

$$D_2(\rho\|\tau_q^{\otimes n}) \leq (1-\epsilon-\log(2(1-q)))n, \quad (32)$$

the probability that virtual cooling or distillation succeeds is bounded by $2^{-\epsilon n}$.

That is, for more general fixed points, the virtual cooling or distillation will succeed with exponentially small probability if the relative entropy has decayed by a factor of $\log(2q)$. As was the case with the concentration
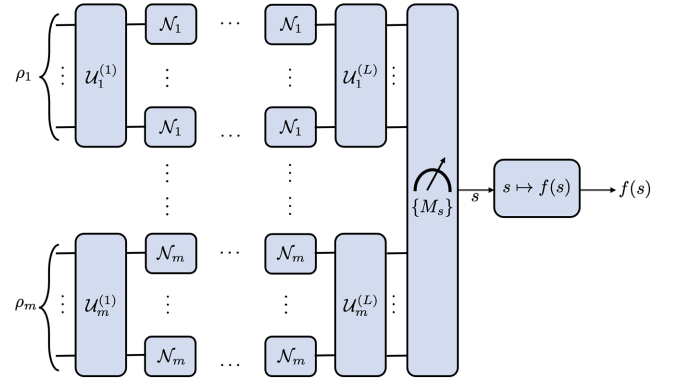

FIG. 1. Schematic of an error-mitigation protocol.

bounds, we see that our bounds become weaker as the fixed point becomes purer. Furthermore, it is also possible to immediately apply the results derived in Appendix C to estimate when the entropy has contracted enough such that the success probability becomes exponentially small.

B. Weak error mitigation with regular estimators

We now see that the techniques of the last sections also readily apply to weak error-mitigation techniques that are regular in a sense that will be made precise later. To the best of our knowledge, all weak error-mitigation techniques have the following basic building blocks and parts.

(1) Take the outcome of m (noisy) quantum circuits $\mathcal{E}_1, \ldots, \mathcal{E}_m$ with initial states $\rho_1, \ldots, \rho_m$.
(2) Add auxiliary qubits and perform a collective noisy circuit Φ on the output of the m circuits.
(3) Perform a measurement on the m systems.
(4) Postprocess the outcomes of the measurements and output an estimate.

This is illustrated in Fig. 1. It is easy to see that points (2), (3), and (4) can all be collectively modeled by applying a global projective measurement $\mathcal{M} := \{M_s\}_{s \in \mathcal{S}}$ on state $\bigotimes_{i=1}^m \mathcal{E}_i(\rho_i) \otimes |0\rangle\langle 0|^{\otimes k}$, where we assume that we have access to k auxiliary systems. Here the projection-valued measure (PVM) is indexed from some classical sample space $\mathcal{S}$, followed by a classical procedure mapping each measured output $s \in \mathcal{S}$ to a real value $f(s)$ through a function $f : \mathcal{S} \to \mathbb{R}$. The hope is then that $f(s)$ provides a good estimate for some property of the noiseless circuit.

Equivalently, we are interested in the probabilistic properties of the observable

$$X := \sum_{s \in \mathcal{S}} f(s)\mathrm{tr}_A(I_S \otimes |0\rangle\langle 0|^{\otimes k} M_s) \quad (33)$$

in the output state $\rho_{\text{out}} := \bigotimes_{i=1}^m \mathcal{E}_i(\rho_i)$ of the original noisy circuit, where we have traced out the auxiliary systems used in the mitigation process.

In order to obtain concentration inequalities for error-mitigation protocols, we impose a bit more structure on the estimators. To make our motivation for our further assumptions clear, we use as our guiding example the most naive of all error-mitigation protocols for an optimization task: sampling m times from the quantum device, evaluating the energy of each outcome and outputting the minimum; i.e., just repeating the experiment often enough. First, we assume that the PVM is indexed by labels $s \in \mathbb{R}^m$. In the case of the minimum strategy discussed before, the individual entries of this vector would correspond to the energy we observed on each one of the m copies. Furthermore, we assume that f is L_f Lipschitz with respect to the ℓ_∞ norm on $\mathbb{R}^m$, i.e.,

$$\sup_{s,s'\in\mathbb{R}^m} \frac{|f(s) - f(s')|}{\|s - s'\|_{\ell_\infty}} \leq L_f.$$

In the case of the minimum strategy, f would correspond to the minimum in $\mathbb{R}^m$, for which we have $L_f = 1$. Let us justify the assumption that f is Lipschitz by looking in a bit more detail at the case of the POVM measuring copies independently. In that case, f being Lipschitz with respect to ℓ_∞ corresponds to requiring that the error-mitigated estimate should not depend too strongly on any individual sample, a robustness condition that is desirable in the presence of noise. Finally, we assume that the error-mitigation procedure concentrates when given trivial, product states:

$$\mathbb{P}_{\sigma^{\otimes m}}(\|s - \mathbb{E}(s)\|_{\ell_\infty} \geq rn) \leq K(m)\exp\left(-\frac{cr^2 n}{\ell_0^2}\right) \quad (34)$$

for some function $K(m)$.

Let us discuss this assumption once again in the case of taking the minimum of measuring the energy of a Hamiltonian H m times. In that case, each measurement satisfies Gaussian concentration for some c and $\ell_0 = \|\sigma^{-1/2} H \sigma^{1/2}\|_L$. Thus, it follows from a union bound that, for taking independent measurements, Eq. (34) holds with $K(m) = m$.

Now that we have formulated the error-mitigation protocol in this way, we can immediately apply the same reasoning as in Proposition IV.1 to understand the concentration properties of the error-mitigation procedure.

Theorem V.1. *For an error-mitigation observable X as in Eq. (33), assume that, for a given state σ, Eq. (34) holds for some function $K(m)$. Furthermore, assume that, for $r, \epsilon > 0$, given that we have, for all $1 \leq i \leq m$, $D_2(\mathcal{E}_i(\rho)\|\sigma) \leq$*

$c(r^2 - \epsilon)n/ml_0^2$, *then*

$$\mathbb{P}_{\rho_{\text{out}}}(|X - f(\mathbb{E}_{\sigma^{\otimes m}}(s))\mathbb{I}| > rL_f n) \leq \exp\left(-\frac{c\epsilon n}{\ell_0^2}\right). \tag{35}$$

We leave the proof of Theorem V.1 to Appendix I. We see that the amount by which the Rényi entropy has to decrease to ensure that we are in the regime where we obtain concentration from Eq. (35) is connected to the Lipschitz constant of X and the number of copies m. For instance, under local depolarizing noise with depolarizing probability p, this happens at depth $\mathcal{O}(p^{-1} \log(ml_0))$.

One way of interpreting the bound in Eq. (35) is that the probability that the estimate we obtain from the output of the error-mitigation algorithm with input given by the noisy states to that with the fixed point of the noise as input is exponentially small. Thus, the noisy outputs are useless: we could have just sampled from the product state $\sigma^{\otimes m}$ instead and observed similar outcomes.

However, it might be hard to control the Lipschitz constant L_f in general scenarios. Moreover, many mitigation protocols in the literature [66–69] involve estimating the mean of random variables that take exponentially large values. Thus, their Lipschitz constant will typically also be exponentially large, constraining the applicability of Theorem V.1.

VI. EXAMPLE: FINDING THE GROUND STATE OF ISING HAMILTONIANS IN THE NISQ ERA

Given a matrix $A \in \mathbb{R}^{n \times n}$ and a vector $b \in \mathbb{R}^n$, we define the Hamiltonian

$$H_I = -\sum_{i,j=1}^n A_{i,j} Z_i Z_j - \sum_{i=1}^n b_i Z_i. \tag{36}$$

It is well known how to formulate various NP-complete combinatorial optimization problems as finding a string that minimizes the energy of H_I. This has motivated the pursuit of NISQ algorithms for this task, including the QAOA [11] or the closely related quantum annealing algorithm.

Let us briefly describe the QAOA algorithm. Given a $P \in \mathbb{N}$ and vectors of parameters $\gamma, \tau \in \mathbb{R}^P$, the QAOA unitary is given by

$$V_{\gamma,\beta} = \prod_{k=1}^P e^{i\beta_k H_X} e^{i\gamma_k H_I}, \tag{37}$$

where $H_X = -\sum_{i=1}^n X_i$. The hope of QAOA is that, by optimizing over parameters γ, β, measuring $V_{\gamma,\beta} |+\rangle^{\otimes n}$ in the computational basis will yield low-energy strings for Hamiltonian (36) even for moderate values of P. In

what follows we distinguish the depth of the QAOA ansatz (denoted by P) from the physical depth of the circuit being implemented in the device (denoted by L).

In recent years, several works have identified limitations on the performance of constant depth circuits in outperforming classical algorithms for this problem [16,18], even in the absence of noise. These results were then later extended to short-time quantum annealing [19].

Taking the noise into consideration, recent works have shown that QAOA is outperformed by efficient classical algorithms at a depth that is proportional to the local noise rate [20]. However, those works only considered the expected value of the output string. Considering that the goal of QAOA is to obtain one low-energy string, to completely discard exponential advantages of QAOA and other related algorithms at a depth that depends only on local noise rates, it is important to also obtain concentration inequalities for the outputs.

As mentioned before, Proposition IV.1 already allows us to conclude that quantum advantage will be lost against classical algorithms at constant depth. With the techniques presented in this work, it is also straightforward to obtain concentration bounds for concrete instances. Indeed, given that a classical algorithm found a string with given energy $-a_C n$, we can easily bound the depth at which the bound in Proposition IV.1 kicks in and the quantum device is exponentially unlikely to yield a better result.

A. Max-cut problem

In this subsection, we analyze the performances of quantum circuits for the max-cut problem. Let $G = (V, E)$ be a graph. The *cut* of a bipartition of V is the number of edges that connect the two parts. The max-cut problem consists in finding the maximum cut of G, which we denote with $C_{\max}$. The best classical algorithm for the max-cut problem is due to Goemans and Williamson [70] and can obtain a string whose cut is at least $0.878\,C_{\max}$. As in Ref. [16], we consider circuits that commute with $\sigma_x^{\otimes n}$, which include the QAOA circuit. We prove that the algorithm by Goemans and Williamson cannot be outperformed by

(i) noiseless circuits with shallow depth (Theorem VI.1);
(ii) noisy circuits with any depth (Theorem VI.2).

We assume that G is bipartite, i.e., $C_{\max} = |E|$, and is regular with degree D, i.e., each vertex belongs to exactly D edges. Without loss of generality, we assume that $V = [n]$. We associate to each bipartition $V = V_0 \cup V_1$, $V_0 \cap V_1 = \emptyset$ the bit string $x \in \{0,1\}^n$ such that $x_i = j$ if $i \in V_j$. We denote with $C(x)$ the cut of such bipartition. We also assume that G satisfies

$$C(x) \geq h \min\{|x|, n - |x|\}, \qquad h = \frac{D}{2} - \sqrt{D-1}, \quad (38)$$

for any $x \in \{0,1\}^n$, where $|x|$ denotes the Hamming weight of x, i.e., the number of components of x that are equal to 1. For any $D \geq 3$, Ramanujan expander graphs constitute an example of graphs with such property [71–73]. Moreover, random D-regular bipartite graphs approach bound (38) with high probability [74].

The max-cut problem for G is equivalent to maximizing the n-qubit Hamiltonian

$$H = \sum_{x \in \{0,1\}^n} C(x)\, |x\rangle\langle x| = \frac{1}{2} \sum_{(j,k) \in E} (\mathbb{I} - \sigma_z^j \sigma_z^k), \quad (39)$$

where, for any $j \in [n]$, σ_z^j is the Pauli-Z matrix acting on qubit j.

Theorem VI.1 (**Noiseless max-cut problem**). *Let G be a regular bipartite graph with n vertices satisfying bound (38), and let H be the associated max-cut Hamiltonian (39). Let ρ be the output of a noiseless quantum circuit as in Definition II.1 made by L layers, where each layer consists of a set of unitary gates acting on mutually disjoint couples of qubits. We assume that the input state of the circuit and each unitary gate commute with $\sigma_x^{\otimes n}$. Then, if*

$$\mathrm{tr}[\rho H] \geq C_{\max}\left(\frac{5}{6} + \frac{\sqrt{D-1}}{3D}\right), \quad (40)$$

we must have

$$L \geq \frac{1}{2} \log_2 \frac{n}{576}. \quad (41)$$

Furthermore, if ρ is generated by the QAOA circuit (37) with depth P, we must have

$$P \geq \frac{1}{2 \log(D+1)} \log \frac{n}{576}. \quad (42)$$

Remark VI.1: For any $D \geq 55$, we have

$$\frac{5}{6} + \frac{\sqrt{D-1}}{3D} < 0.878; \quad (43)$$

therefore, any quantum algorithm that outperforms the algorithm by Goemans and Williamson must generate a state satisfying Eq. (40).

Remark VI.2: Under the same hypotheses of Theorem VI.1, Bravyi *et al.* [16, Corollary 1] proved that

$$P \geq \frac{1}{3(D+1)} \log_2 \frac{n}{4096}. \quad (44)$$

Our result (42) provides an exponential improvement over Eq. (44) with respect to D. Already for $D = 55$, the right-hand side of Eq. (44) is larger than 1 only for $n = \Omega(10^{54})$,

while the right-hand side of Eq. (42) is larger than 1 already for $n = \Omega(10^6)$.

Proof of Theorem VI.1. Circuit made of two-qubit gates. From Proposition III.2, ρ satisfies a $(2, \infty)$-Poincaré inequality with constant

$$C \leq 2^{2L+2}. \tag{45}$$

Let

$$
\begin{aligned}
A &= \left\{ x \in \{0,1\}^n : d_H(x, x_{\text{opt}}) \leq \frac{n}{3} \right\}, \\
B &= \left\{ x \in \{0,1\}^n : d_H(x, \bar{x}_{\text{opt}}) \leq \frac{n}{3} \right\},
\end{aligned} \tag{46}
$$

and let X be the random outcome obtained measuring ρ in the computational basis. Proposition H.1 in Appendix H implies that

$$\mathbb{P}(X \in A) = \mathbb{P}(X \in B) \geq \tfrac{1}{4}, \tag{47}$$

and Eq. (12) of Theorem III.1 together with Eq. (45) implies that

$$\frac{n}{3} \leq d_H(A, B) \leq 2^{L+3}\sqrt{n}. \tag{48}$$

Claim (41) follows.

QAOA circuit. From Proposition III.2, ρ satisfies a $(2, \infty)$-Poincaré inequality with constant

$$C \leq 4(D+1)^{2P}. \tag{49}$$

Proceeding as in the previous case, we get

$$\frac{n}{3} \leq d_H(A, B) \leq 8(D+1)^P\sqrt{n}, \tag{50}$$

and claim (42) follows. ∎

Theorem VI.2 (Noisy max-cut problem). *Under the same hypotheses as Theorem VI.1, let each layer of the circuit be followed by depolarizing noise with depolarizing probability p applied to each qubit. Then,*

$$n \leq 3 \cdot 2^{2/p+8}. \tag{51}$$

For $p = 0.1$, bound (51) gives $n \leq 8 \times 10^8$.

Proof. If $n \leq 3072$, bound (51) is empty. We can then assume that $n \geq 3072$. Proceeding as in the proof of

Theorem VI.1 and employing Proposition III.1 in place of Proposition III.2, we get

$$n \leq 3 \cdot 2^{2L+8}. \tag{52}$$

Let us consider the following operator associated with the Hamming distance from x_{opt}:

$$K = \sum_{x \in \{0,1\}^n} d_H(x, x_{\text{opt}}) |x\rangle\langle x| - \frac{n}{2}\, \mathbb{I}. \tag{53}$$

We have $\operatorname{tr} K = 0$ and $\|K\|_L = 1$; therefore, Proposition IV.1 implies that, for any $\epsilon > 0$, upon measuring K on ρ we have

$$\mathbb{P}(|K| \geq (\epsilon + (1-p)^{2L})n) \leq 2e^{-\epsilon n/2}. \tag{54}$$

Proposition H.1 in Appendix H implies that

$$\mathbb{P}\left(\left|\frac{1}{n}d_H(X, x_{\text{opt}}) - \frac{1}{2}\right| \geq \frac{1}{6}\right) \geq \frac{1}{2}, \tag{55}$$

and choosing in Eq. (54)

$$\epsilon = \tfrac{1}{6} - (1-p)^{2L}, \tag{56}$$

we get

$$(1-p)^{2L} \geq \tfrac{1}{6} - \tfrac{4}{n}\ln 2 \geq e^{-2}; \tag{57}$$

hence,

$$L \leq -\frac{1}{\ln(1-p)} \leq \frac{1}{p}. \tag{58}$$

The claim follows by combining Eqs. (52) and (58). ∎

B. Short-time evolution of local Hamiltonians

Quantum annealing constitutes another family of heuristic algorithms to solve optimization problems. Similar to the variational algorithms discussed earlier, the goal in quantum annealing is to find the lowest energy of a classical Hamiltonian that encodes the optimization problem. To find the lowest energy of the optimization Hamiltonian, we can start from a local Hamiltonian whose ground state is easy to prepare, for example $-\sum_i X_i$, and continuously change the Hamiltonian to the desired optimization Hamiltonian H_I:

$$H(t) = -a(t)\sum_i X_i + b(t)H_I \tag{59}$$

with $a(0) = b(T) = 1$, $a(T) = b(0) = 0$, and T the final evolution time. The adiabatic theorem [75] guarantees that

if we start from the ground state of the initial Hamiltonian and evolve the system slowly enough, the final state would be close to the ground state of the optimization Hamiltonian, which can be found by measurement in the computational basis at the final time. Since noise restricts the total time that coherence in the system is preserved, understanding the limitations of short-time evolution of local Hamiltonians seems crucial. The presented $(2, \infty)$-Poincaré inequality provides bounds on the performance of short-time quantum annealers.

Proposition VI.1 (**Short-time evolution of local Hamiltonians**). *Let σ be the quantum state generated by evolving a product state with a continuous-time local quantum process as in Sec. III B for time $t \geq 0$. Let μ_σ be the probability distribution of the outcome of the measurement in the computational basis performed on σ. Then, for any $A, B \subseteq \{0,1\}^V$, we have*

$$ d_H(A, B) \leq \sqrt{c_0 + c_1 e^{vt}} \sqrt{|V|} (\mu_\sigma(A)^{-1/2} + \mu_\sigma(B)^{-1/2}), $$

where d_H denotes the Hamming distance, $v = eb(2D - 1)$, D is the maximum degree of the interaction graph, b is the maximum interaction strength and

$$ c_0 = 64M\delta^\delta, \qquad c_1 = \frac{64M}{2D - 1} Li_{-2(\delta - 1)}(e^{-1}), \quad (60) $$

where $Li_s(z)$ is the polylogarithm function of order s and argument z and δ is the spatial dimension of the interaction graph.

Remark VI.3: Crucially, both c_0 and c_1 are independent of the number of qubits.

Proof of Proposition VI.1. We start by deriving an upper bound on C_t of Theorem III.1. We note that by the definition of i_0 we have $2\delta - 1 \leq d(i_0) \leq 2\delta$, and therefore, using $|S_e(k)| \leq M k^{\delta - 1}$, we have $i_0 \leq \sum_{k=0}^{2\delta} |S_e(k)| \leq 2\delta M \delta^{\delta - 1} + 1 = 2M\delta^\delta + 1$. Also, we have

$$ \sum_{i=i_0}^{n} d(i)^{\delta-1} e^{-d(i)} \leq \sum_{k=2\delta-1}^{n} |S_e(k)| k^{\delta-1} e^{-k} $$

$$ \leq M \sum_{k=2\delta-1}^{n} k^{2(\delta-1)} e^{-k} $$

$$ \leq M Li_{-2(\delta-1)}(e^{-1}). \quad (61) $$

Putting these two bounds together, we have

$$ C_t \leq (c_0 + c_1 e^{vt})^2. \quad (62) $$

The claim follows by applying Theorem III.1. ∎

Considering the example of generating a generalized GHZ state, where $d_H(A, B) = n$ and $\mu_\sigma(A) = \mu_\sigma(B) = 1/2$, we have

$$ \frac{1}{v} \log \left(\frac{n}{8c_1} - \frac{c_0}{c_1} \right) \leq t, \quad (63) $$

and, therefore, at least $O(\log(n))$ time is required to generate generalized GHZ states using local Hamiltonians. Note that this bound also provides a minimum time required by local Hamiltonians to simulate unitaries that are capable of generating generalized GHZ states starting from product states, such as n-qubit fan-out gates.

The short-time evolution of local Hamiltonians also limits their performance to solve the max-cut problem discussed in Sec. VI A. Note that both the initial state and the annealing Hamiltonian of Eq. (59) with the final Hamiltonian H_I corresponding to the max-cut problem commute with $\sigma_x^{\otimes n}$, and therefore the techniques of Theorem VI.1 directly lead to a proof for the limitation of short-time evolution of the local Hamiltonian for the optimization task.

Proposition VI.2. *Consider the max-cut problem Hamiltonian H_I as discussed in Theorem VI.1, and the corresponding annealing Hamiltonian in the form (59). Let ρ be evolved states after time T. Then, if*

$$ \mathrm{tr}[\rho H] \geq C_{\max} \left(\frac{5}{6} + \frac{\sqrt{D-1}}{3D} \right), \quad (64) $$

we must have

$$ T \geq \frac{1}{v} \ln \left(\frac{\sqrt{n}}{12c_1} - \frac{c_0}{c_1} \right). \quad (65) $$

Proof. From Eqs. (48) and (12) of Theorem III.1, we have

$$ \frac{n}{3} \leq d_H(A, B) \leq 4\sqrt{C_t}\sqrt{n}, \quad (66) $$

which can be combined with Eq. (62) to obtain

$$ \frac{\sqrt{n}}{12} \leq c_0 + c_1 e^{vT}. \quad (67) $$

The claim follows. ∎

C. Noisy QAOA beyond unital noise

In this subsection we discuss the performance of our bounds for QAOA beyond the case of unital noise. As mentioned before, if the noise is not unital, our bounds on the relative entropy decay are not independent of the circuit being implemented. Thus, we need to pick a promising family of QAOA parameters to apply our results.
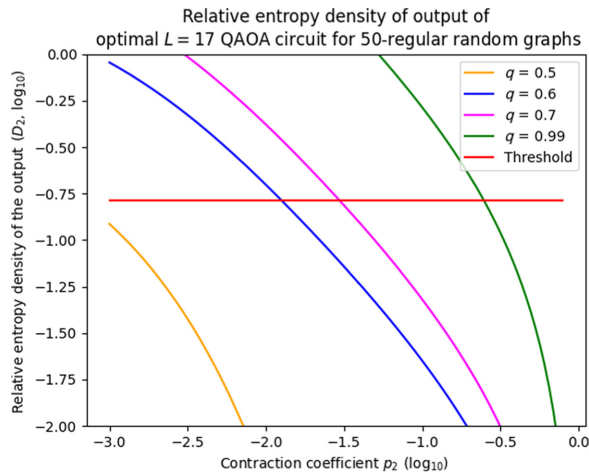
FIG. 2. Relative entropy density of the output of a QAOA circuit of $P = 17$ layers for various fixed points $\tau_q^{\otimes n}$ as a function of the contraction coefficient and $D = 50$. We use the optimal parameters found in Ref. [76] for our circuit. The threshold we use is that in Eq. (68) and we use Corollary C.1 to estimate the relative entropy decay. Although we see that our bounds have a worse performance as $q \to 1$, the amount of noise we can tolerate is still independent of the system's size.

A natural candidate of instances to analyze is the max-cut problem on random regular graphs of high girth. This is because in Ref. [76] the authors derived the optimal parameters for QAOA for such graphs in the large-n limit for up to 17 layers. Furthermore, they showed that these QAOA circuits achieve an expected value for the cut that is higher than what known provably efficient classical algorithms achieve. Although these parameters are only optimal in the absence of noise, we analyze their performance in the presence of nonunital noise driving the system to the classical state $\tau_q^{\otimes n}$ with $\tau_q = q |0\rangle \langle 0| + (1 - q) |1\rangle \langle 1|$.

As explained in Appendix C 3, we show that, as long as output ρ of a noisy QAOA circuit satisfies

$$D_2(\rho \| \tau_q^{\otimes n}) < \frac{((1 - 2q)^2 D/2 + 2\sqrt{D}/\pi)^2}{2D^2} n \qquad (68)$$

for a D-regular graph, the probability that the noisy circuit outperforms classical methods is exponentially small. See Fig. 2 for when this is achieved in terms of the contraction coefficient.

Although Fig. 2 seems to suggest that advantage is only lost at high noise levels as the fixed point becomes purer, recall that, when implementing the QAOA circuit on the actual device, the circuit depth will be significantly larger than 17. Indeed, in the plot, we take $D = 50$, which means that a circuit of depth at least 50 of two-qubit gates is required to implement each layer of $e^{i\gamma_i H_I t}$. If we further incorporate the compilation of gates and the fact that NISQ devices are unlikely to have all-to-all connectivity, which imposes extra layers of SWAP gates, the depth required to

implement each layer of QAOA with $D = 50$ will conservatively be of order at least 10^2. Thus, it is also reasonable to assume that the effective noise rate when implementing a layer of the QAOA circuit will be 2 orders of magnitude larger than the physical noise rate.

More generally, our bounds predict that quantum advantage will be lost whenever the QAOA parameters satisfy $\beta_k \to 0$ as $k \to \infty$. This is the case for the optimal parameters found in Ref. [76]. This is because, for such parameters, the relative entropy between the output of the circuit and $\tau_q^{\otimes n}$ decays to 0. This is illustrated more clearly in the continuous-time case of quantum annealing we discuss now.

D. Noisy quantum annealing beyond unital noise

In this subsection, we illustrate the bound in Proposition IV.2 for the case of noisy annealers with a linear schedule. That is, function f in the statement is just given by $f(t) = (1 - t)$. Furthermore, we assume that the time-independent Lindlbadian of spectral gap 1 is driving the system to the product state $\tau_q^{\otimes q}$ with $\tau_q = q |0\rangle \langle 0| + (1 - q) |1\rangle \langle 1|$ for $q < 1/2$.

Proposition VI.3. *For $0 < q \le 1/2$ and $T > 0$, let*

$$r_2 = 2 \frac{1 - q}{\log(q^{-1})} \qquad (69)$$

and let

$$h(T) = e^{-r_2 T} \log \left(\frac{1 + 2(q - q^2)^{1/2}}{4(q - q^2)} \right)$$
$$+ \frac{(2q - 1)(1 - e^{-r_2 T} rT - e^{-r_2 T})}{(q(1 - q))^{1/2} r_2^2 T}. \qquad (70)$$

Furthermore, let $\mathcal{T}_t$ be defined as in Proposition IV.2 and $f(t) = (1 - t)$. Then, for the initial state $|+\rangle^{\otimes n}$ and $\rho_T = \mathcal{T}_T(|+\rangle \langle +|^{\otimes n})$, we have

$$\mathbb{P}_{\rho_T}(H_I \le (\mathrm{tr}([\tau_q^{\otimes n} H_I] - 2^{-1/2}(h(T) + \epsilon)^{1/2} \|H_I\|_{\mathrm{Lip}} n) \mathbb{I})$$
$$\le \exp\left(-\frac{\epsilon n}{2}\right). \qquad (71)$$

We refer the reader to Appendix C 4 for a discussion of this result and Proposition C.3 in the same section for a proof. But the take-away message from Proposition VI.3 is that we can still derive concentration inequalities beyond unital noise. However, the bounds get looser as $q \to 0$ (i.e., the fixed point becomes pure) and the decay of the relative entropy is polynomial instead of exponential.

We can reach similar conclusions for the purity of the output and, thus, for the probability that virtual cooling succeeds.

Proposition VI.4. *For $0 < q \leq 1/2$ and $T > 0$, let*

$$r_2 = 2\frac{1-q}{\log(q^{-1})}, \tag{72}$$

and let $h(T)$ be as in Eq. (70). Furthermore, let $\mathcal{T}_t$ be defined as in Proposition C.1 in Appendix C and $f(t) = (1-t)$. For the initial state $|+\rangle^{\otimes n}$, let T be large enough for $h(T) \leq 1 - \log(2(1-q)) - \epsilon$ to hold for some $\epsilon > 0$. Then the probability that virtual cooling or distillation succeeds is at most $\exp(-\epsilon n)$.

We refer the reader to Appendix C 4 for a proof.

VII. CONCLUSION AND OPEN PROBLEMS

In this work we have used techniques of quantum optimal transport to derive various concentration inequalities for quantum circuits. In particular, we showed quadratic concentration for shallow circuits and Gaussian concentration for noisy circuits at large enough depth and Lipschitz observables.

By applying such inequalities to variational quantum algorithms such as QAOA or quantum annealing algorithms, we showed that, for most instances, the probability that these algorithms outperform classical algorithms is exponentially small whenever the circuit has a nontrivial density of errors. Furthermore, we obtained self-contained and simplified proofs of previous results on the limitations of QAOA.

Our work demonstrates the relevance of quantum optimal transport methods to near-term quantum computing. Furthermore, it closes a few important gaps in previous results on limitations of variational quantum algorithms.

An important problem that is left by our work is whether it is also possible to obtain Gaussian concentration inequalities for the outputs of shallow circuits. After posting the first version of the present work, Anshu and Metger [77] found a different method based on polynomial approximations for showing that the output distributions in fact satisfy a stronger Gaussian concentration bound, hence answering this question.

ACKNOWLEDGMENTS

APPENDIX A: NOTATION

We consider a set V corresponding to a system of $|V| = n$ qudits, and denote by $\mathcal{H}_V = \bigotimes_{v \in V} \mathbb{C}^d$ the Hilbert space of n qudits and by $\mathcal{B}(\mathcal{H}_V)$ the algebra of linear operators on $\mathcal{H}_V$. We denote by $\mathcal{O}_V$ the self-adjoint linear operators on $\mathcal{H}_V$, whereas $\mathcal{O}_V^T \subset \mathcal{O}_V$ is the subspace of traceless self-adjoint linear operators. By $\mathcal{O}_V^+$ we denote the subset of positive semidefinite linear operators on $\mathcal{H}_V$, and $\mathcal{S}_V \subset \mathcal{O}_V^+$ denotes the set of quantum states. Similarly, we denote by $\mathcal{P}_V$ the set of probability measures on $[d]^V$. Given an operator $X \in \mathcal{B}(\mathcal{H}_V)$, we denote by $X^\dagger$ its adjoint with respect to the inner product of $\mathcal{H}_V$. Similarly, the adjoint of a linear map $\mathcal{N} : \mathcal{B}(\mathcal{H}_V) \to \mathcal{B}(\mathcal{H}_V)$ with respect to the trace inner product is denoted by $\mathcal{N}^\dagger$. For any subset $A \subseteq V$, we use the standard notation $\mathcal{O}_A, \mathcal{S}_A, \ldots$ for the corresponding objects defined on subsystem A. Given a state $\rho \in \mathcal{S}_V$, we denote by ρ_A its marginal on subsystem A. For any $X \in \mathcal{O}_V$, we denote by $\|X\|_p$ its Schatten p norm. For any region $A \subset V$, the identity on $\mathcal{O}_A$ is denoted by $\mathbb{I}_A$, or more simply $\mathbb{I}$. Given an observable O, we define $\langle O \rangle_\sigma = \text{tr}[\sigma O]$. Moreover, given a number $a \in \mathbb{R}$, we define $\{O \geq a\}$ to be the projector onto the subspace spanned by the eigenvectors of O corresponding to eigenvalues greater than or equal to a. We denote the probability of measuring an eigenvalue of O greater than $a \in \mathbb{R}$ in state σ as $\mathbb{P}_\sigma(O \geq a) := \text{tr}[\sigma \{O \geq a\}]$. Given two probability measures μ, ν over a common measurable space, $\mu \ll \nu$ means that μ is absolutely continuous with respect to ν. We make use of the sandwiched Rényi divergences [41,42] of order $\alpha \in (1, +\infty)$. For two states ρ, σ such that the support of ρ is included in the support of σ, they are defined as

$$D_\alpha(\rho \| \sigma) = \frac{1}{\alpha - 1} \log \text{tr}[(\sigma^{(1-\alpha)/2\alpha} \rho \sigma^{(1-\alpha)/2\alpha})^\alpha].$$

We also consider the relative entropy we obtain by taking the limit $\alpha \to \infty$,

$$D_\infty(\rho \| \sigma) = \log(\|\sigma^{-1/2} \rho \sigma^{-1/2}\|_\infty),$$

and the usual Umegaki relative entropy between two quantum states ρ, σ, defined as

$$D(\rho \| \sigma) := \text{tr}[\rho (\log \rho - \log \sigma)],$$

which corresponds to the limit $\alpha \to 1$. In case the support of ρ is not contained in that of σ, all the divergences above are defined to be $+\infty$.

APPENDIX B: RÉNYI DIVERGENCES AND CONCENTRATION INEQUALITIES

In this section, we show how to use Rényi divergences to transfer results about concentration from one state to another. These divergences can be used to transfer concentration inequalities between states as follows.

Lemma B.1 (Transferring concentration inequalities). *Let ρ and $\sigma > 0$ be two quantum states on $\mathcal{H}_V$. Then, for any positive operator-valued measure (POVM) element $0 \le E \le \mathbb{I}$ and $\alpha > 1$, we have*

$$\mathrm{tr}[E\rho] \le \exp\left[\frac{\alpha-1}{\alpha}(D_\alpha(\rho\|\sigma) + \log(\mathrm{tr}[E\sigma]))\right]. \quad \text{(B1)}$$

In particular, if σ satisfies the Gaussian concentration inequality

$$\mathbb{P}_\sigma\big(|O - \langle O\rangle_\sigma| \ge a|V|\big) \le K\, \exp\left(-\frac{ca^2|V|}{\|\sigma^{-1/2}O\sigma^{1/2}\|_L^2}\right)$$

for some constants $c, K > 0$, then, for any $\alpha > 1$,

$$\mathbb{P}_\rho(|O - \langle O\rangle_\sigma| \ge a|V|)$$
$$\le \exp\left[\frac{\alpha-1}{\alpha}\left(D_\alpha(\rho\|\sigma) - \frac{ca^2|V|}{\|\sigma^{-1/2}O\sigma^{1/2}\|_L^2} + \log(K)\right)\right]. \quad \text{(B2)}$$

Proof. We have

$$\mathrm{tr}[E\rho] = \mathrm{tr}[\sigma^{-(1-\alpha)/2\alpha}E\sigma^{-(1-\alpha)/2\alpha}\sigma^{(1-\alpha)/2\alpha}\rho\sigma^{(1-\alpha)/2\alpha}]$$
$$\le \mathrm{tr}[(\sigma^{-(1-\alpha)/2\alpha}E\sigma^{-(1-\alpha)/2\alpha})^{\alpha'}]^{1/\alpha'}$$
$$\times \mathrm{tr}[(\sigma^{(1-\alpha)/2\alpha}\rho\sigma^{(1-\alpha)/2\alpha})^\alpha]^{1/\alpha}$$

by an application of Hölder's inequality with α' the Hölder conjugate of α. Next, by the Araki-Lieb-Thirring inequality,

$$\mathrm{tr}[(\sigma^{-(1-\alpha)/2\alpha}E\sigma^{-(1-\alpha)/2\alpha})^{\alpha'}]$$
$$\le \mathrm{tr}[\sigma^{-(1-\alpha)\alpha'/2\alpha}E^{\alpha'}\sigma^{-(1-\alpha)\alpha'/2\alpha}]$$
$$\le \mathrm{tr}[\sigma^{-(1-\alpha)\alpha'/2\alpha}E\sigma^{-(1-\alpha)\alpha'/2\alpha}], \quad \text{(B3)}$$

where in the last inequality we have used the facts that $\alpha' > 1$ and $E \le \mathbb{I}$. Furthermore, as α' is the Hölder conjugate of

α, we have $1/\alpha' = (\alpha - 1)/\alpha$ and then

$$\mathrm{tr}[\sigma^{-(1-\alpha)\alpha'/2\alpha}E\sigma^{-(1-\alpha)\alpha'/2\alpha}] = \mathrm{tr}[\sigma E].$$

The claim in Eq. (B1) then follows from a simple manipulation and by noting that

$$\mathrm{tr}[(\sigma^{(1-\alpha)/2\alpha}\rho\sigma^{(1-\alpha)/2\alpha})^\alpha]^{1/\alpha} = \exp\left(\frac{\alpha-1}{\alpha}D_\alpha(\rho\|\sigma)\right).$$

Equation (B2) also immediately follows from plugging in the Gaussian concentration bound. ∎

APPENDIX C: ENTROPIC CONVERGENCE RESULTS

In this section we collect some results that allow us to estimate the sandwiched Rényi divergence between the output of a noisy quantum circuit or annealer and the fixed point of the noise affecting the device. In essence, these results are a generalization of the results of Ref. [20, Lemma 1 and Theorem 1]. In that work, the authors showed precisely the same bounds as here, but only for the Umegaki relative entropy. However, their proofs can immediately be adapted to our setting with Rényi divergences. Thus, we restrict ourselves to showing how to obtain a convergence result for discrete-time circuits and do not describe the same proof for continuous time in full detail.

Lemma C.1 (Lemma 1 of Ref. [20]). *Let $\mathcal{N} : \mathcal{B}(\mathcal{H}_V) \to \mathcal{B}(\mathcal{H}_V)$ be a quantum channel with unique fixed point $\sigma > 0$ that satisfies a strong data-processing inequality with constant $p_\alpha > 0$ for some $\alpha > 1$. That is,*

$$D_\alpha(\mathcal{N}(\rho)\|\sigma) \le (1 - p_\alpha)D_\alpha(\rho\|\sigma) \quad \text{(C1)}$$

for all states ρ. Then, for any other quantum channels $\Phi_1, \ldots, \Phi_m : \mathcal{B}(\mathcal{H}_V) \to \mathcal{B}(\mathcal{H}_V)$, we have

$$D_\alpha\left(\prod_{t=1}^m (\Phi_t \circ \mathcal{N})(\rho)\bigg\|\sigma\right) \le (1 - p_\alpha)^m D_\alpha(\rho\|\sigma)$$
$$+ \sum_{t=1}^m (1 - p_\alpha)^{m-t} D_\infty(\Phi_t(\sigma)\|\sigma). \quad \text{(C2)}$$

Proof. For $m = 1$, this follows from the data-processed triangle inequality of Ref. [78, Theorem 3.1]. In their notation, it states that, for any quantum channel P, states

ρ, σ, σ', and $\alpha \geq 1$, we have

$$D_\alpha(P(\rho)\|\sigma) \leq D(\rho\|\sigma') + D_\infty(P(\sigma')\|\sigma).$$

Setting $P = \Phi_1$ and $\sigma' = \sigma$, in their notation it implies that

$$D_\alpha((\Phi_1 \circ \mathcal{N})(\rho)\|\sigma) \leq D_\alpha(\mathcal{N}(\rho)\|\sigma) + D_\infty(\Phi_1(\sigma)\|\sigma)$$
$$\leq (1 - p_\alpha)D_\alpha(\rho\|\sigma) + D_\infty(\Phi_1(\sigma)\|\sigma). \quad (C3)$$

Let us now assume the claim to be true for some $m = k$. Then, for $m = k + 1$, we have

$$D_\alpha\left(\prod_{t=1}^{k+1}(\Phi_t \circ \mathcal{N})(\rho)\middle\|\sigma\right) \leq (1 - p_\alpha)^k D_\alpha$$

$$((\Phi_{k+1} \circ \mathcal{N})(\rho)\|\sigma) + \sum_{t=1}^{k}(1 - p_\alpha)^{k-t}D_\infty(\Phi_t(\sigma)\|\sigma) \quad (C4)$$

by our induction hypothesis. Applying Eq. (C3) to the first term in Eq. (C4), the strong data-processing inequality, we obtain the claim. ∎

Note that Lemma C.1 implies that the Rényi divergence will converge to 0 whenever $\Phi_t(\sigma) \simeq \sigma$ as $t \to \infty$. This is always the case for unitary circuits under unital noise, as the fixed point is the maximally mixed state and is invariant under unitaries, but is also expected to hold for QAOA circuits. See Sec. VI for examples of such circuits.

We can also show similar statements for continuous-time evolutions under noise to also study quantum simulators or annealers.

Lemma C.2 (Theorem 1 of Ref. [20]). *Let $\mathcal{L} : \mathcal{B}(\mathcal{H}_V) \to \mathcal{B}(\mathcal{H}_V)$ be a Lindbladian with fixed point σ. Suppose that, for some $\alpha > 1$, we have, for all $t > 0$ and initial states, an $r_\alpha > 0$ such that*

$$D_\alpha(e^{t\mathcal{L}}(\rho)\|\sigma) \leq e^{-r_\alpha t}D_\alpha(\rho\|\sigma). \quad (C5)$$

Moreover, let $\mathcal{H}_t : \mathcal{B}(\mathcal{H}_V) \to \mathcal{B}(\mathcal{H}_V)$ be given by $\mathcal{H}_t(X) = i[X, H_t]$ for some time-dependent Hamiltonian H_t. Moreover, let $\mathcal{T}_t$ be the evolution of the system under the Lindbladian $\mathcal{S}_t = \mathcal{L} + \mathcal{H}_t$ from time 0 to t. Then, for all states ρ and times $t > 0$,

$$D_\alpha(\mathcal{T}_t(\rho)\|\sigma) \leq e^{-r_\alpha t}D_\alpha(\rho\|\sigma)$$

$$+ \int_0^t e^{-r_\alpha(t-\tau)}\|\sigma^{-1/2}[\sigma, H_\tau]\sigma^{-1/2}\|_\infty \, d\tau. \quad (C6)$$

Thus, armed with contraction inequalities like those in Eqs. (C1) and (C5), it is straightforward to obtain estimates on Rényi entropies. For completeness, we collect some known results and techniques to obtain such contraction inequalities in the next section.

1. Contraction results for sandwiched Rényi divergences

Let us now collect some known results to obtain inequalities like those in Eqs. (C1) and (C5). We focus on the case where the noise has a product form, i.e., $\mathcal{N} = \bigotimes_{i=1}^{n} \mathcal{N}_i$, where $\mathcal{N}_i$ acts only on qubit i. Although it is straightforward to generalize the results to the case in which there is a different channel acting on each qubit, we make the simplifying assumption that all local channels are the same. Furthermore, we focus on inequalities that tensorize. This means that q_α will not scale with the size of the system n. To the best of our knowledge, strong data-processing inequalities are not available for Rényi entropies beyond product channels.

Let us start with the continuous-time setting, as more is known there. For continuous time, the contraction of Rényi entropies was systematically studied in Ref. [63]. In particular, in Theorem 4.3 of Ref. [63] the authors related bounds on the optimal decay rate r_α to so-called logarithmic Sobolev inequalities [59,62,64,79]. It is beyond the scope of this article to review logarithmic Sobolev inequalities and we focus instead on the contraction rate these tools give to the problem at hand.

If we have a Lindbladian of the form

$$\mathcal{L}^{(n)} = \mathcal{L} \otimes \mathrm{id}_{n-1} + \mathrm{id}_1 \otimes \mathcal{L} \otimes \mathrm{id}_{n-2} + \cdots + \mathrm{id}_{n-1} \otimes \mathcal{L}$$

with unique fixed point $\bigotimes_{i=1}^{n} \sigma$, then

$$D_2\left(e^{t\mathcal{L}^{(n)}}(\rho)\middle\|\bigotimes_{i=1}^{n} \sigma\right) \leq e^{-r_2 t}D_2\left(\rho\middle\|\bigotimes_{i=1}^{n} \sigma\right) \quad (C7)$$

holds with

$$r_2 = 2\lambda(\mathcal{L})\frac{1 - 1/\|\sigma^{-1}\|}{\log(\|\sigma^{-1}\|)}, \quad (C8)$$

where $\lambda(\mathcal{L})$ is the spectral gap of the local Linbladian $\mathcal{L}$. For instance, for generalized depolarizing noise, we have $\lambda(\mathcal{L}) = 1$. The take-home message of Eq. (C8) is that, as long as $\|\sigma^{-1}\| = \mathcal{O}(1)$, the rate with which the sandwiched Rényi-2 divergence contracts is constant as well. It is also possible to use similar tools to derive the contraction for other values of $\alpha > 1$ and we refer the reader to Refs. [61, 63] for a more detailed discussion. However, to the best of our knowledge, all known results exhibit a similar scaling to that in Eq. (C8) and we do not discuss this further.

In discrete time, the best results available are, to the best of our knowledge, those of Ref. [61, Corollary 5.5, 5.6]. To parse their results, we first need to introduce some notation. For a given σ, we denote by $\Gamma_\sigma^\alpha : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ the map $X \mapsto \sigma^{\alpha/2}X\sigma^{\alpha/2}$ and by $\mathcal{D}_{p,\sigma}$ the generalized depolarizing channel converging to state σ [i.e., $\rho \mapsto (1-p)\rho + p\sigma$].

It follows from Corollary 5.6 of Ref. [61] that if, for a quantum channel $\mathcal{N}_i$ with fixed point σ, we have

$$\|\Gamma^{-1/2} \circ \mathcal{N}_i \circ \mathcal{D}_{p,\sigma}^{-1} \circ \Gamma^{1/2}\|_{2 \to 2} \leq 1, \qquad \text{(C9)}$$

then, for any state ρ on n qudits,

$$D_2\left(\bigotimes_{i=1}^n \mathcal{N}_i(\rho) \middle\| \bigotimes_{i=1}^n \sigma\right)$$

$$\leq (1-p)^{\|\sigma^{-1}\|_\infty - 1/\|\sigma^{-1}\|_\infty \log(\|\sigma^{-1}\|_\infty)} D_2\left(\rho \middle\| \bigotimes_{i=1}^n \sigma\right). \tag{C10}$$

The expressions in Eqs. (C9) and (C10) may seem daunting at first, so let us digest them a bit further and summarize their message. First, note that Eq. (C9) only involves one copy of the quantum channel, whereas the expression in Eq. (C10) involves arbitrarily many. Thus, this is an example of an inequality that tensorizes. Furthermore, note that Eq. (C9) can be verified efficiently. This is because it just corresponds to checking whether the operator norm of a linear operator is smaller than or equal to one or not, which can be computed in polynomial time. Thus, by performing a binary search on the values of p for which the inequality holds, we can approximate the largest p for which it holds. Then Eq. (C10) tells us that once we establish such an inequality, the Rényi-2 divergence will contract by a rate that is independent of the system size. The take-home message of Eq. (C10) is essentially the same as that of Eq. (C7). As long as $\|\sigma^{-1}\|_\infty = \mathcal{O}(1)$, the Rényi-2 divergence will contract with a constant rate. This corresponds to the setting in which each local fixed point does not have a purity scaling with system size.

2. Specializing Lemma C.1 to QAOA and quantum annealing

In the main text we only considered quantum circuits that are affected by unital noise. The reason for that is that then one can use Lemma C.1 to obtain the exponential decay of the relative entropy to the maximally mixed state independently of the circuit that is being implemented.

However, it is still possible to obtain closed formulas for the relative entropy decay for QAOA-like circuits, as we show now. We are still going to depart from the assumption that the noise affecting the device has a product state $\sigma_q = \bigotimes_{i=1}^n \tau_q$ as its fixed point, with

$$\tau_q = q |0\rangle \langle 0| + (1-q) |1\rangle \langle 1| \tag{C11}$$

for some $q \in [0,1]$.

Recall that, for H_I, the Ising Hamiltonian whose energy we wish to minimize, $H_X = -\sum_i X_i$, and $\gamma, \beta \in \mathbb{R}^P$, the

QAOA unitary is given by

$$V_{\gamma,\beta} = \prod_{k=1}^P e^{i\beta_k H_X} e^{i\gamma_k H_I}. \tag{C12}$$

In order to obtain an estimate of the relative entropy decay under a noisy version of this circuit, we need to analyze the expressions

$$D_\infty(e^{i\gamma_k H_I} \sigma_q e^{-i\gamma_k H_I} \| \sigma_q), \qquad D_\infty(e^{i\beta_k H_X} \sigma_q e^{-i\beta_k H_X}) \| \sigma_q).$$

We then have the following result.

Lemma C.3. *Let $\beta, \gamma \in \mathbb{R}^P$ be given and, for $q \in [0,1]$, σ_q as in Eq. (C11). Moreover, for β_k, q, define $z(\beta_k, q)$ as*

$$z(\beta_k, q) = 2\cos(2\beta_k) + \frac{\sin^2(\beta_k)}{q(1-q)}.$$

Then:

$$D_\infty(e^{i\beta_k H_X} \sigma_q e^{-i\beta_k H_X} \| \sigma_q)$$

$$= n \log\left(\frac{z(\beta_k, q) + \sqrt{z(\beta_k, q)^2 - 4}}{2}\right). \tag{C13}$$

Proof. As both $e^{i\beta_k H_X}$ and σ are of tensor product form, we obtain by the additivity of the max relative entropy

$$D_\infty(e^{i\beta_k H_X} \sigma e^{-i\beta_k H_X}) \| \sigma) = n D_\infty(e^{-i\beta_k X} \tau_q e^{i\beta_k H_X} \| \tau_q).$$

A simple yet tedious computation shows that

$$\|\tau_q^{-1/2} e^{-i\beta_k X} \tau_q e^{i\beta_k H_X} \tau_q^{-1/2}\| = \frac{z(\beta_k, q) + \sqrt{z(\beta_k, q)^2 - 4}}{2}. \tag{C14}$$

Taking the logarithm yields the claim. ∎

Before we state the entropy decay we obtain for QAOA circuits, let us briefly comment on the scaling of Eq. (C13). First, note that either in the limit $q \to 1/2$ or $\beta_k \to 0$, we have the right-hand side of Eq. (C13). The first case corresponds to the fixed point being the maximally mixed state, but the second corresponds to mixer unitaries for which the total time evolution is small.

On the other hand, if we let $q \to 0$ or $q \to 1$, then we see that the right-hand side of Eq. (C13) goes to infinity. We then have the following result.

Corollary C.1 (Relative entropy decay for QAOA). *Let $\beta, \gamma \in \mathbb{R}^P$ be given, and let τ_q and z be defined as before.*

Moreover, let $\mathcal{N}$ be such that

$$D_\alpha(\mathcal{N}(\rho)\|\sigma_q) \le (1-p_\alpha)D_\alpha(\rho\|\sigma_q). \qquad \text{(C15)}$$

Then, for any initial state ρ, we have

$$D_\alpha\left(\prod_{k=1}^{P}(e^{i\beta_k H_I} \circ \mathcal{N} \circ e^{i\beta_k H_X} \circ \mathcal{N})(\rho)\middle\|\sigma_q\right)$$

$$\le (1-p_\alpha)^{2P}D_\alpha(\rho\|\sigma_q) + \sum_{k=1}^{P}(1-p_\alpha)^{2(P-k)}n\log$$

$$\times\left(\frac{z(\beta_k,q)+\sqrt{z(\beta_k,q)^2-4}}{2}\right). \qquad \text{(C16)}$$

Furthermore, for the case of $\rho = |+\rangle\langle+|^{\otimes n}$ and $\alpha = 2$, we have

$$D_2(|+\rangle\langle+|^{\otimes n}\|\sigma_q) = n\log$$

$$\left(\frac{q^{-1}+(1-q)^{-1}+2(q(1-q))^{-1/2}}{4}\right). \qquad \text{(C17)}$$

Proof. The first step is to observe that $D_\infty(e^{i\gamma_k H_I}\sigma_q e^{-i\gamma_k H_I})\|\sigma) = 0$. This follows from the fact that $e^{i\beta_k H_I}$ is a diagonal unitary and, thus, commutes with σ_q. The claim then follows from combining Lemma C.1 and the result of Lemma C.3. To obtain the expression in Eq. (C17), note that D_2 tensorizes and the two underlying states are products. Thus, we only need to compute $D_2(|+\rangle\langle+|\|\tau_q)$, a simple computation. ∎

From our previous discussion, it is straightforward to identify the conditions under which Eq. (C16) converges to 0 as $P \to \infty$. First, the case $q = 1/2$, which corresponds to unital noise and we already covered at length in the main text. Second, whenever we have $\beta_k \to 0$ as $k \to \infty$. This is because the relative entropy terms in Eq. (C16) at depth k are suppressed by $(1-p_\alpha)^{2(P-k)}$. Thus, only at depths $k \simeq P$ is the relative entropy not suppressed.

Interestingly, parameters β, γ for which QAOA is expected to perform well fulfill this condition [80,81]. To see this, it is fruitful to interpret QAOA as a trotterized version of quantum annealing, where we start with Hamiltonian H_X and adiabatically modify it to H_I. It is then clear that at late times of the computation, the Hamiltonian will approximate H_I and the fixed point of the noise will be approximately preserved by the unitary evolution.

We can make this precise by deriving the analogous version of Corollary C.1 for quantum annealing.

Proposition C.1. *Let $\mathcal{L} : \mathcal{B}(\mathcal{H}_V) \to \mathcal{B}(\mathcal{H}_V)$ be a Lindbladian with fixed point σ_q defined as before with $q \ge 1/2$.*

Suppose that, for some $\alpha > 1$, we have, for all $t > 0$ and initial states, an $r_\alpha > 0$ such that

$$D_\alpha(e^{t\mathcal{L}}(\rho)\|\sigma) \le e^{-r_\alpha t}D_\alpha(\rho\|\sigma). \qquad \text{(C18)}$$

Moreover, for functions $f,g : [0,1] \to \mathbb{R}$ and $T > 0$, let $\mathcal{H}_t : \mathcal{B}(\mathcal{H}_V) \to \mathcal{B}(\mathcal{H}_V)$ be given by $\mathcal{H}_t(X) = i[X, f(t/t)H_X + g(t/T)H_I]$. Let $\mathcal{T}_t$ be the evolution of the system under the Lindbladian $\mathcal{S}_t = \mathcal{L} + \mathcal{H}_t$ from time 0 to $t \le T$. Then, for all states ρ,

$$D_\alpha(\mathcal{T}_T(\rho)\|\sigma) \le e^{-r_\alpha T}D_\alpha(\rho\|\sigma) + 2ne^{-r_\alpha T}$$

$$\left(\sqrt{\frac{p}{1-p}} - \sqrt{\frac{1-q}{q}}\right)\int_0^T e^{r_\alpha t}|f(t/T)|\,dt. \qquad \text{(C19)}$$

Proof. From Lemma C.2 we see that all we need to obtain the claim is to estimate

$$\int_0^T e^{-r_\alpha(T-t)}\|\sigma_q^{-1/2}[\sigma_q,H_t]\sigma_q^{-1/2}\|_\infty\,dt.$$

As before, because $[H_I,\sigma_q] = 0$, this simplifies to

$$\int_0^T e^{-r_\alpha(T-t)}\|\sigma_q^{-1/2}[\sigma_q,H_\tau]\sigma_q^{-1/2}\|_\infty\,dt$$

$$= \int_0^T e^{-r_\alpha(T-t)}|f(t/T)|\|\sigma_q^{-1/2}[\sigma_q,H_X]\sigma_q^{-1/2}\|_\infty\,dt$$

$$\le ne^{-r_\alpha T}\int_0^T e^{-r_\alpha t}|f(t/T)|\|\tau_q^{-1/2}[\tau_q,X]\tau_q^{-1/2}\|_\infty\,dt,$$

where in the last step we applied a triangle inequality using $H_X = -\sum_i X_i$ and the fact that $\sigma_q = \bigotimes_{i=1}^n \tau_q$. The claim follows after noting that

$$\|\tau_q^{-1/2}[\tau_q,X]\tau_q^{-1/2}\|_\infty = \left(\sqrt{\frac{p}{1-p}} - \sqrt{\frac{1-p}{p}}\right).$$

This completes the proof. ∎

As adiabatic theorems require that $f(1) = 0$ to make sure that we observe a good overlap with the ground state [75], it follows that the Rényi entropy will typically decay to 0 even under nonunital noise for quantum annealers. However, note once again that our bounds perform poorly whenever the fixed point is close to pure and whenever function f does not decay fast enough to 0 around 1.

3. QAOA and quantum annealing on random regular graphs of high girth

In the previous section we established estimates on the relative entropy decay of QAOA circuits (Corollary C.1) and quantum annealers (Proposition C.1) under nonunital

noise. Such estimates can then be combined with Theorem IV.1 to obtain concentration inequalities for the outputs of these circuits. One important caveat is that Corollary C.1 and Proposition C.1 depend on the actual circuit being implemented. Thus, we cannot give universal bounds on the performance of such circuits that depend only on the depth and the noise level as was the case for unital noise.

However, Corollary C.1 can still be readily applied for a given choice of QAOA parameters and we exemplify the performance of the bounds on QAOA on the max-cut of random D-regular graphs under noise. The motivation to study this particular class of instances is many. First, the asymptotic value of both the ground-state energy and that of the standard semidefinite program relaxation are known. It is known [82,83] that, for the Ising model on a random D-regular graph on n nodes, the ground-state energy density scales like

$$-\Pi_* \sqrt{D} + o(\sqrt{D}) \qquad (C20)$$

with $\Pi_* = 0.763\,166\ldots$ the Parisi constant. The value that assumption-free efficient classical algorithms [84] achieve is given by $-2/\pi \sqrt{D}$ with $2/\pi \simeq 0.6366$. The fact that these values are known makes it straightforward to analyze at which energies the output of a noisy quantum algorithm will be outperformed by efficient classical algorithms.

Furthermore, there is a natural choice for the value of the QAOA parameters to pick for the circuit. Indeed, in Ref. [76] the authors computed the optimal parameters for QAOA on such graphs for depths up to $P = 17$. Note, however, that these are the optimal values as the system's size goes to infinity and in the absence of noise. Nevertheless, they provide a good testing ground for our bounds.

To start our analysis, note that if we define the one-qubit state $\tau_q = q \,|0\rangle\langle 0| + (1 - q)\,|1\rangle\langle 1|$ as before and let $H_{I,D}$ be the Ising Hamiltonian on a D-regular graph, then we have

$$\mathrm{tr}(H_{I,D}\tau_q^{\otimes n}) = (1 - 2q)^2 \frac{nD}{2}. \qquad (C21)$$

To see this, note that the expectation value of each $Z_i Z_j$ term will be $(1 - 2q)^2$ and the graph is assumed to be D regular. We then obtain Eq. (C21) by noting that there are $nD/2$ edges in the graph. As the expected value of the energy achieved by classical algorithms is $-2n\sqrt{D}/\pi$, quantum advantage is lost if we deviate by less than $(1 - 2q)^2 nD/2 + 2n\sqrt{D}/\pi$ from the mean under $\tau_q^{\otimes n}$.

We then have the following result.

Proposition C.2. *Let ρ be a quantum state on n qubits and assume that, for some $q \in (0, 1)$, $\epsilon > 0$, and $D > 0$,*

we have

$$D_2(\rho \| \tau_q^{\otimes n}) \leq \frac{((1 - 2q)^2 D/2 + 2/\pi\sqrt{D})^2 - \epsilon D^2}{2D^2} n. \qquad (C22)$$

Then the probability that the outcome of measuring ρ in the computational basis provides a lower energy than efficient classical algorithms for max-cut on random D-regular high girth algorithms is at most $e^{-\epsilon n/2}$.

Proof. Note that we have $\|H_{I,D}\|_{\mathrm{Lip}} = D$, as the graph is D regular. By Theorem IV.1 we have

$$\mathbb{P}_\rho\left(\left|H_{I,D} - (1 - 2q)^2 \frac{D}{2}\right| \geq an\right)$$

$$\leq \exp\left(\frac{1}{2}\left(D_2(\rho \| \tau_q^{\otimes n}) - \frac{a^2 n}{2D^2}\right)\right). \qquad (C23)$$

By our previous discussion, we know that we need to deviate from the mean with respect to the state $\tau_q^{\otimes n}$ by at least $(1 - 2q)^2 nD/2 + 2n\sqrt{D}/\pi$ so the quantum algorithm outperforms classical algorithms. Thus, we can pick $a = (1 - 2q)^2 D/2 + 2\sqrt{D}/\pi$ in Eq. (20) as our measure of when advantage is lost. It is then easy to see that, for our bound on D_2 in Eq. (C22), we find that the right-hand side of Eq. (C23) is $e^{-\epsilon n/2}$, which shows the claim. ∎

Proposition C.2 allows us to conclude that if the output ρ of a QAOA circuit satisfies

$$D_2(\rho \| \tau_q^{\otimes n}) < \frac{((1 - 2q)^2 D/2 + 2\sqrt{D}/\pi)^2}{2D^2} n \qquad (C24)$$

then quantum advantage is lost. In Ref. [80, Table 4] the authors gave optimal parameters that in the noiseless case outperform known efficient classical algorithms. We can then insert these parameters into the bound obtained in Corollary C.1 to estimate at which noise levels advantage is lost.

It is important to stress once again that these parameters are only known to be optimal in the absence of noise and in the limit of nodes and degree going to infinity. However, we believe that they still provide a natural choice of parameters to analyze under noise. Importantly, note that Eq. (C24) once again only requires the relative entropy to contract by a constant factor before advantage is lost as long as $q \neq 1/2$.

In Fig. 2 of the main text we plot the performance of QAOA with the parameters for $P = 17$ as predicted by our bounds. In the absence of noise these QAOA circuits outperform efficient classical algorithms, but we show that this is not necessarily the case in the presence of noise.

Note that the values of γ_i are irrelevant for the analysis. The values of β_i we use are

$$\beta = [0.6375, 0.5197, 0.4697, 0.4499, 0.4255, 0.4054,$$
$$0.3832, 0.3603, 0.3358, 0.3092, 0.2807, 0.2501,$$
$$0.2171, 0.1816, 0.1426, 0.1001, 0.0536]. \quad (C25)$$

4. Computations required for Sec. VI D

In this subsection we collect some auxiliary computations required to arrive at the conclusion of the example discussed in Sec. VI D. Our goal is to evaluate Eq. (C19) for the case where the initial state is given by $|+\rangle = (1/\sqrt{2})(|0\rangle + |1\rangle)$ and the annealing schedule is linear, $f(t) = (1 - t)$. Furthermore, for simplicity, we assume that the local Lindbladians $\mathcal{L}_i$ have as a fixed point the state τ_q for $q \leq 1/2$ and spectral gap $\lambda = 1$. The results can then be easily rescaled to obtain the bounds for other values of the spectral gap.

The first observation we make is that, under these assumptions, Eq. (C8) implies that

$$r_2 \geq 2\frac{1-q}{\log(q^{-1})}. \quad (C26)$$

Furthermore, a simple yet tedious calculation shows that

$$D_2(|+\rangle \langle+|^{\otimes n} \| \tau_q^{\otimes n}) = n \log\left(\frac{1 + 2(q - q^2)^{1/2}}{4(q - q^2)}\right) \quad (C27)$$

and the integral in Eq. (C19) evaluates to

$$\int_0^T e^{r_2 t} |f(t/T)| \, dt = \frac{e^{rT} - rT - 1}{r^2 T}.$$

Putting all of these elements together we obtain the bound

$$n^{-1} D_2(\mathcal{T}_T(\rho) \| \tau_q^{\otimes n}) \leq e^{-r_2 T} \log\left(\frac{1 + 2(q - q^2)^{1/2}}{4(q - q^2)}\right)$$
$$+ \frac{(2q - 1)(1 - e^{-r_2 T} rT - e^{-r_2 T})}{(q(1 - q))^{1/2} r_2^2 T}, \quad (C28)$$

where r_2 is lower bounded in Eq. (C26). Furthermore, by combining Theorem 2 of Ref. [29] and Theorem 7 of Ref. [26] we conclude that, for τ_q and O satisfying $[O, \tau_q^{\otimes n}] = 0$,

$$\mathbb{P}_{\tau_q^{\otimes n}}(|O - \text{tr}[O\tau_q^{\otimes n}]\mathbb{I}| \geq r) \leq 2\exp\left(-\frac{2r^2}{n\|O\|_{\text{Lip}}^2}\right). \quad (C29)$$

The one-sided bound also holds without the prefactor 2. Now that we have a contraction result for the Rényi divergence and a concentration inequality for the fixed point of

the noise, it is straightforward to also obtain concentration bounds for the output of the noisy quantum annealer with Theorem IV.1.

Indeed, we conclude the following.

Proposition C.3. *For* $0 < q \leq 1/2$ *and* $T > 0$, *let*

$$r_2 = 2\frac{1-q}{\log(q^{-1})} \quad (C30)$$

and let

$$h(T) = e^{-r_2 T} \log\left(\frac{1 + 2(q - q^2)^{1/2}}{4(q - q^2)}\right)$$
$$+ \frac{(2q - 1)(1 - e^{-r_2 T} rT - e^{-r_2 T})}{(q(1 - q))^{1/2} r_2^2 T}. \quad (C31)$$

Furthermore, let $\mathcal{T}_t$ *be defined as in Proposition C.1 and* $f(t) = (1 - t)$. *Then, for the initial state* $|+\rangle^{\otimes n}$, *we have*

$$\mathbb{P}_{\mathcal{T}_T(|+\rangle\langle+|^{\otimes n})}(H_I \leq (\text{tr}([\tau_q^{\otimes n} H_I] - 2^{-1/2}(h(T) + \epsilon)^{1/2}$$
$$\times \|H_I\|_{\text{Lip}} n)\mathbb{I}) \leq \exp\left(-\frac{\epsilon n}{2}\right). \quad (C32)$$

Proof. As proved in Eq. (C28), it follows that at time T we have

$$n^{-1} D_2(\mathcal{T}_T(|+\rangle \langle+|^{\otimes n}) \| \tau_q^{\otimes n}) \leq h(T), \quad (C33)$$

and we have the concentration inequality in Eq. (C29) for $\tau_q^{\otimes n}$.

We now pick the parameter $r = 2^{-1/2}\|H_I\|_L(h(T) + \epsilon)^{1/2} n$ for the concentration inequality. It then follows from Theorem IV.1 that

$$\mathbb{P}_{\mathcal{T}_T(|+\rangle\langle+|^{\otimes n})}(H_I \leq (\text{tr}[\tau_q^{\otimes n} H_I] - 2^{-1/2}(h(T) + \epsilon)^{1/2}$$
$$\times \|H_I\|_L n)\mathbb{I}) \leq \exp\left(\frac{n}{2}(h(T) - (h(T) - \epsilon))\right), \quad (C34)$$

which yields the claim. $\blacksquare$

APPENDIX D: BOUNDS ON PURITY AND HIGHER MOMENTS

We now obtain upper bounds on $\text{tr}[\rho^k]$ for ρ the output of a noisy circuit and $k \geq 2$. The motivation for such bounds comes from understanding the success probability of virtual distillation or cooling protocols [33,34]. Roughly speaking, these protocols have as their goal to prepare the quantum state $\rho^k/\text{tr}[\rho^k]$ from k copies of ρ. As explained before in Sec. V A, the success probability of these protocols is $\text{tr}[\rho^k]$. We prove that at constant depth the success probability becomes exponentially small in system size.

The first observation we make is that $k \mapsto \mathrm{tr}[\rho^k]$ is monotonically decreasing in k. Thus, it suffices to show that the purity $\mathrm{tr}[\rho^2]$ is exponentially small at constant depth under noise. We start with the following lemma.

Lemma D.1. *Let* $\tau_q = q|0\rangle\langle0| + (1-q)|1\rangle\langle1|$, *and assume without loss of generality that* $q \leq 1/2$. *Then, for any state* $\rho \in \mathcal{S}_V$ *with* $n = |V|$ *such that*

$$D_2(\rho\|\tau_q^{\otimes n}) \leq (1 - \epsilon - \log(2(1-q)))n, \quad (D1)$$

we have

$$\mathrm{tr}[\rho^2] \leq 2^{-\epsilon n}. \quad (D2)$$

Proof. First note that we have

$$\mathrm{tr}[\rho^2] = 2^{-n+D_2(\rho\|\mathbb{I}/2^n)}. \quad (D3)$$

Thus, the claim follows if we show that Eq. (D1) implies that $D_2(\rho\|\mathbb{I}/2^n) \leq (1-\epsilon)n$. From the data-processed triangle inequality [78, Theorem 3.1] we obtain

$$D_2(\rho\|\mathbb{I}/2^n) \leq D_2(\rho\|\tau_q^{\otimes n}) + D_\infty(\tau_q^{\otimes n}\|\mathbb{I}/2^n)$$
$$= D_2(\rho\|\tau_q^{\otimes n}) + \log(2(1-q))n, \quad (D4)$$

and so it follows from Eq. (D1) that

$$D_2(\rho\|\mathbb{I}/2^n) \leq (1-\epsilon)n,$$

and we obtain the bound by inserting the equation above into Eq. (D3). ∎

As before, in the case of unital noise ($q = 1/2$) we are able to obtain statements that are independent from the circuit being implemented. Moreover, the limitations are even more striking than for our concentration bounds, as summarized in the statement below.

Proposition D.1. *Let* $\mathcal{N}_V$ *be a depth-*L *unitary circuit interspersed by a unital noise channel* $\mathcal{N}$ *such that*

$$D_2\left(\mathcal{N}_V(\rho)\left\|\frac{I}{2^n}\right.\right) \leq (1-r_2)D\left(\rho\left\|\frac{I}{2^n}\right.\right).$$

Then, for any initial state ρ *and* $k \geq 2$,

$$\mathrm{tr}[\mathcal{N}_V(\rho)^k]$$
$$\leq \mathrm{tr}[\mathcal{N}_V(\rho)^2] \leq \exp(-\log(2)(1-(1-r_2)^L)n).$$

Proof. The result immediately follows from combining Lemma D.1 with the fact that the output satisfies

$$D_2\left(\mathcal{N}_V(\rho)\left\|\frac{I}{2^n}\right.\right) \leq (1-r_2)^L n. \quad ∎$$

Thus, we see from Eq. (D5) that, unless $r_2 = \mathcal{O}(n^{-1})$, the purity will already be exponentially small in system size and will accept the resulting state. We conclude that in the case of unital noise, they are only effective for circuits that have a constant number of expected errors.

In the case of $q < 1/2$, the success probability will only be exponential at depths after which the Rényi-2 divergence has contracted by more than $\log(2q)$. We can use results like Proposition C.1 and Corollary C.1 to estimate when this happens. However, in these cases we expect that the bounds we currently have only predict that the depth at which the purity becomes exponential is of order $\mathcal{O}(p_\alpha^{-1})$.

Let us illustrate this more concretely with noisy annealers. For the case of a linear schedule, under the same conditions as for Proposition C.3, we obtain the following result.

Proposition D.2. *For* $0 < q \leq 1/2$ *and* $T > 0$, *let*

$$r_2 = 2\frac{1-q}{\log(q^{-1})} \quad (D5)$$

and let

$$h(T) = e^{-r_2 T}\log\left(\frac{1 + 2(q-q^2)^{1/2}}{4(q-q^2)}\right)$$
$$+ \frac{(2q-1)(1 - e^{-r_2 T}rT - e^{-r_2 T})}{(q(1-q))^{1/2}r_2^2 T}. \quad (D6)$$

Furthermore, let $\mathcal{T}_t$ *be defined as in Proposition C.1 and* $f(t) = (1-t)$. *For the initial state* $|+\rangle^{\otimes n}$, *let* T *be large enough for* $h(T) \leq 1 - \log(2(1-q)) - \epsilon$ *to hold for some* $\epsilon > 0$. *Then,*

$$\mathrm{tr}[\mathcal{T}_T(|+\rangle\langle+|^{\otimes n})^k] \leq \mathrm{tr}[\mathcal{T}_T(|+\rangle\langle+|^{\otimes n})^2] \leq \exp(-\epsilon n).$$

Proof. This statement immediately follows from Lemma D.1 and the fact that $n^{-1}D_2(\mathcal{T}_T(|+\rangle\langle+|^{\otimes n})\|\tau_q^{\otimes n}) \leq h(T)$, as we showed in Eq. (C28). ∎

As explained in the main text, such bounds can be applied to bound the probability that virtual distillation protocols work.

APPENDIX E: PROOF OF THEOREM III.1

In this section, we prove the consequences of the $(2, \infty)$-Poincaré inequality stated in Theorem III.1, which we restate below for clarity of exposition.

Theorem E.1. *Assume that state* $\sigma \in \mathcal{S}_V$ *satisfies a* $(2, \infty)$-*Poincaré inequality with constant* $C > 0$. *Then the following statements hold.*

(i) Noncommutative transport-variance inequality: for any two states $\rho_1, \rho_2 \in \mathcal{S}_V$ with corresponding densities $X_j := \sigma^{-1/2} \rho_j \sigma^{-1/2}$,

$$W_1(\rho_1, \rho_2) \leq \sqrt{C|V|}(\|X_1 - \mathbb{I}\|_\sigma + \|X_2 - \mathbb{I}\|_\sigma). \tag{E1}$$

(ii) Measured transport-variance inequality: denote by $\mu_\sigma \in \mathcal{P}_V$ the probability measure induced by the measurement of σ in the computational basis. Then, for any $\nu \ll \mu_\sigma$,

$$W_1(\nu, \mu_\sigma) \leq \sqrt{C|V| \operatorname{Var}_{\mu_\sigma}(d\nu/d\mu_\sigma)}.$$

Moreover, for any two sets $A, B \subset [d]^V$, their Hamming distance $d_H(A, B)$ satisfies the following symmetric concentration inequality:

$$d_H(A, B) \leq \sqrt{C|V|}(\mu_\sigma(A)^{-1/2} + \mu_\sigma(B)^{-1/2}). \tag{E2}$$

(iii) Concentration of observables: for any observable $O \in \mathcal{O}_V$ and $r > 0$,

$$\mathbb{P}_\sigma\left(|O - \langle O \rangle_\sigma| \geq r\right) \leq \frac{C|V| \|O\|_L^2}{r^2}. \tag{E3}$$

Proof. (i) By the Cauchy-Schwarz inequality, we have, for $X = \sigma^{-1/2} \rho \sigma^{-1/2}$,

$$\begin{aligned}
W_1(\rho, \sigma) &= \sup_{\|H\|_L \leq 1} \operatorname{tr}[(\rho - \sigma)H] \\
&= \sup_{\|H\|_L \leq 1} \langle X - \mathbb{I}, H - \mathbb{I} \rangle_\sigma \\
&\leq \sup_{\|H\|_L \leq 1} \|X - \mathbb{I}\|_\sigma \|H - \mathbb{I}\|_\sigma \\
&\leq \sup_{\|H\|_L \leq 1} \|X - \mathbb{I}\|_\sigma \operatorname{Var}_\sigma(H)^{\frac{1}{2}} \\
&\leq \sqrt{C|V|}\|X - \mathbb{I}\|_\sigma, \tag{E4}
\end{aligned}$$

where Eq. (E4) follows from Eq. (10). Therefore, by the triangle inequality, for any two states ρ_1, ρ_2 with corresponding densities $X_j = \sigma^{-1/2} \rho_j \sigma^{-1/2}$,

$$W_1(\rho_1, \rho_2) \leq \sqrt{C|V|}(\|X_1 - \mathbb{I}\|_\sigma + \|X_2 - \mathbb{I}\|_\sigma).$$

(ii) Since the quantum variance $\operatorname{Var}_\sigma(O_F)$ and the classical variance $\operatorname{Var}_{\mu_\sigma}(F)$ coincide for a classical Lipschitz function F with $O_F := \sum_{\epsilon \in [d]^V} F(\epsilon)|\epsilon\rangle\langle\epsilon|$, we have

$$\operatorname{Var}_{\mu_\sigma}(F) \leq C|V|\|F\|_L^2, \tag{E5}$$

where we have further used the fact that classical and quantum Lipschitz constants coincide, i.e., $\|F\|_L = \|O_F\|_L$; see

Ref. [29, Proposition 7]. Next, by the Cauchy-Schwarz inequality, we have, for all $\nu \ll \mu_\sigma$ and $g := d\nu/d\mu_\sigma$,

$$\begin{aligned}
W_1(\nu, \mu_\sigma) &= \sup_{\|F\|_L \leq 1} \nu(F) - \mu_\sigma(F) \\
&= \sup_{\|F\|_L \leq 1} \mu_\sigma((g - 1)(F - 1)) \\
&\leq \sup_{\|F\|_L \leq 1} \|g - 1\|_{L_2(\mu_\sigma)} \|F - 1\|_{L_2(\mu_\sigma)} \\
&= \sup_{\|F\|_L \leq 1} (\operatorname{Var}_{\mu_\sigma}(g) \operatorname{Var}_{\mu_\sigma}(F))^{1/2} \\
&\leq \sqrt{C|V| \operatorname{Var}_{\mu_\sigma}(g)}.
\end{aligned}$$

The proof of Eq. (E2) is standard [44]. Respectively denote by ν_A and ν_B the probability measures

$$\nu_A(C) := \frac{\mu_\sigma(A \cap C)}{\mu_\sigma(A)}, \qquad \nu_B(C) := \frac{\mu_\sigma(B \cap C)}{\mu_\sigma(B)}. \tag{E6}$$

Then, by the dual formulation of the Wasserstein distance in terms of couplings, we have

$$\begin{aligned}
d_H(A, B) &\leq W_1(\nu_A, \nu_B) \\
&\leq W_1(\nu_A, \mu_\sigma) + W_1(\nu_B, \mu_\sigma) \\
&\leq \sqrt{C|V|}(\operatorname{Var}_{\mu_\sigma}(d\nu_A/d\mu_\sigma)^{1/2} \\
&\quad + \operatorname{Var}_{\mu_\sigma}(d\nu_B/d\mu_\sigma)^{1/2}) \\
&= \sqrt{C|V|}(\mu_\sigma(A)^{-1/2} + \mu_\sigma(B)^{-1/2}),
\end{aligned}$$

where the last line follows from bounding the variance

$$\begin{aligned}
\operatorname{Var}_{\mu_\sigma}(d\nu_A/d\mu_\sigma) &= \int \left(\frac{d\nu_A}{d\mu_\sigma}\right)^2 d\mu_\sigma - 1 \\
&= \int \frac{1_A(x)}{\mu_\sigma(A)^2} d\mu_\sigma(x) - 1 \leq \frac{1}{\mu_\sigma(A)}, \tag{E7}
\end{aligned}$$

and similarly for $\operatorname{Var}_{\mu_\sigma}(d\nu_B/d\mu_\sigma)$.

Statement (iii) is a direct consequence of the $(2, \infty)$-Poincaré inequality and Chebyshev's inequality. ∎

APPENDIX F: CONTROLLING THE LIPSCHITZ CONSTANT

In this appendix, we derive the bounds on the Lipschitz constant of observables evolving according to a local continuous- and discrete-time evolution, namely Proposition III.1, Proposition III.2, and Proposition III.3. We start by proving the $(2, \infty)$-Poincaré inequality for product states. In fact, we prove a slight refinement of it. Let us start by defining, for any $O \in \mathcal{O}_V$ and any $v \in V$,

$$\partial_v O := 2 \min\{\|O - \mathbb{I}_v \otimes O_{v^c}\|_\infty : O_{v^c} \in \mathcal{O}_{v^c}\}.$$

By definition, we hence have $\|O\|_L = \max_{v \in V} \partial_v O$.

Lemma F.1. *For any product state $\rho \in \mathcal{S}_V$ and all $O \in \mathcal{O}_V$,*

$$\mathrm{Var}_\rho(O) \leq \sum_{v \in V} (\partial_v O)^2 \leq |V| \|O\|_L^2.$$

Proof. We fix an arbitrary ordering $\{1, \ldots, n\}$ of the vertices V. For any $i \in [n]$, let $O_{i^c} \in \mathcal{O}_{i^c}$ satisfy

$$\partial_i O = 2\|O - \mathbb{I}_i \otimes O_{i^c}\|_\infty.$$

Given a subregion $A \subseteq V$, we define $\langle O \rangle_{\rho_A} := \mathrm{tr}_A[\rho_A O] \otimes \mathbb{I}_A$. Then, by a telescopic sum argument,

$$\mathrm{Var}_\rho(O) = \mathrm{tr}\left[\rho\left(\sum_{i=1}^n \langle O \rangle_{\rho_{1\ldots i-1}} - \langle O \rangle_{\rho_{1\ldots i}}\right)^2\right]$$

$$= \sum_{i,j=1}^n \mathrm{tr}[\rho(\langle O \rangle_{\rho_{1\ldots i-1}} - \langle O \rangle_{\rho_{1\ldots i}})$$

$$\times (\langle O \rangle_{\rho_{1\ldots j-1}} - \langle O \rangle_{\rho_{1\ldots j}})]$$

$$\overset{(1)}{=} \sum_{i=1}^n \mathrm{tr}[\rho\,(\langle O \rangle_{\rho_{1\ldots i-1}} - \langle O \rangle_{\rho_{1\ldots i}})^2]$$

$$= \sum_{i=1}^n \mathrm{tr}[\rho(\langle O - \mathbb{I}_i \otimes O_{i^c} \rangle_{\rho_{1\ldots i-1}} - \langle O$$

$$- \mathbb{I}_i \otimes O_{i^c} \rangle_{\rho_{1\ldots i}})^2]$$

$$\leq \sum_{i=1}^n \|\langle O - \mathbb{I}_i \otimes O_{i^c} \rangle_{\rho_{1\ldots i-1}}$$

$$- \langle O - \mathbb{I}_i \otimes O_{i^c} \rangle_{\rho_{1\ldots i}}\|_\infty^2$$

$$\leq 4 \sum_{i=1}^n \|O - \mathbb{I}_i \otimes O_{i^c}\|_\infty^2$$

$$= \sum_{i=1}^n (\partial_i O)^2.$$

In equality (1) above, we used the orthogonality relation that, for any $i \neq j$,

$$\mathrm{tr}[\rho(\langle O \rangle_{\rho_{1\ldots i-1}} - \langle O \rangle_{\rho_{1\ldots i}})(\langle O \rangle_{\rho_{1\ldots j-1}} - \langle O \rangle_{\rho_{1\ldots j}})] = 0,$$
(F1)

since $\rho \equiv \bigotimes_{v \in V} \rho_v$ is assumed to be a tensor product. ∎

We need the following technical result in what follows.

Proposition F.1. *Let $v \in V$, and let I_v be the future light cone of v with respect to the quantum channel $\mathcal{N}_V$ on $\mathcal{S}_V$. Then, for any $O \in \mathcal{O}_v$,*

$$\partial_v \mathcal{N}_V^\dagger(O) \leq 2 \sum_{w \in I_v} \partial_w O \leq 2|I_v| \|O\|_L.$$

Proof. For any $w \in V$, let $O_{w^c} \in \mathcal{O}_{w^c}$ such that

$$\partial_w O = 2\|O - \mathbb{I}_w \otimes O_{w^c}\|_\infty.$$

Let $|I_v| = k$, and let us label with $\{1, \ldots, k\}$ the elements of I_v. The observable $\mathcal{N}^\dagger(\mathbb{I}_{I_v} \otimes \mathrm{tr}_{I_v} O/d^{|I_v|})$ does not act on v; therefore,

$$\partial_v \mathcal{N}^\dagger(O) \leq 2\left\|\mathcal{N}^\dagger\left(O - \mathbb{I}_{I_v} \otimes \frac{\mathrm{tr}_{I_v} O}{d^{|I_v|}}\right)\right\|_\infty$$

$$\leq 2\left\|O - \mathbb{I}_{I_v} \otimes \frac{\mathrm{tr}_{I_v} O}{d^{|I_v|}}\right\|_\infty$$

$$\leq 2\sum_{i=1}^k \left\|\mathbb{I}_{1\ldots i-1} \otimes \frac{1}{d^{i-1}} \mathrm{tr}_{1\ldots i-1} O\right.$$

$$\left. - \mathbb{I}_{1\ldots i} \otimes \frac{1}{d^i} \mathrm{tr}_{1\ldots i} O\right\|_\infty$$

$$= 2\sum_{i=1}^k \left\|\mathbb{I}_{1\ldots i-1} \otimes \frac{1}{d^{i-1}} \mathrm{tr}_{1\ldots i-1}[O - \mathbb{I}_i \otimes O_{i^c}]\right.$$

$$\left. - \mathbb{I}_{1\ldots i} \otimes \frac{1}{d^i} \mathrm{tr}_{1\ldots i}[O - \mathbb{I}_i \otimes O_{i^c}]\right\|_\infty$$

$$\leq 4\sum_{i=1}^k \|O - \mathbb{I}_i \otimes O_{i^c}\|_\infty$$

$$= 2\sum_{i=1}^k \partial_i O.$$

This completes the proof. ∎

Next, we consider the noisy circuit introduced in Eq. (9). For any noisy gate $\mathcal{N}_{\ell,e}$, we denote by $\sigma_{\ell,e'}$ the environment state of copy e' of set e, and by $\mathcal{U}_{\ell,\{e,e'\}}$ the unitary dilation of $\mathcal{N}_{\ell,e}$ acting on set e and its copy e', so that

$$\mathcal{N}_{\ell,e}(\rho) = \mathrm{tr}_{e'}(\mathcal{U}_{\ell,\{e,e'\}}(\rho \otimes \sigma_{\ell,e'})).$$

We also denote by $\mathcal{U}_{VA}$ the composition of the tensor products of dilations $\mathcal{U}_{\ell,\{e,e'\}}$, where system A represents the total environment resulting from all the dilations previously defined. In other words, defining $\sigma_A := \bigotimes_{\ell,e} \sigma_{\ell,e'}$, we have

$$\mathcal{N}_V(\rho) = \mathrm{tr}_A[\mathcal{U}_{VA}(\rho \otimes \sigma_A)] \equiv \rho_{\mathrm{out}}.$$

We denote by $I_{\mathcal{U}_{VA}}$ the light cone of $\mathcal{U}_{VA}$ with respect to the decomposition

$$\mathcal{H}_V \otimes \mathcal{H}_A \equiv \bigotimes_{v \in V} \mathcal{H}_v \otimes \bigotimes_{\ell,e} \mathcal{H}_{\ell,e'}.$$
(F2)

Proposition F.2. *For any $O \in \mathcal{O}_V$ and any product state $\rho \in \mathcal{S}_V$ with $\rho_{\text{out}} := \mathcal{N}_V(\rho)$,*

$$\text{Var}_{\rho_{\text{out}}}(O) \leq 4 \|O\|_L^2$$
$$\times \left(|V| I_{\mathcal{N}_V}^2 + \max_\ell |E_\ell| \sum_{\ell=1}^L \max_{e \in E_\ell} I(e, L - \ell)^2 \right),$$

where, given a set $e \in E_\ell$ and $m \in \mathbb{N}$, $I(e, L - \ell)$ denotes the set of all vertices in V in the light cone of set e for the circuit constituted of the last $L - \ell$ layers of $\mathcal{N}_V$.

Remark F.1: In the noiseless setting where there are no ancilla systems, by a closer look into the proof below, we can get rid of the sum over layers and hence recover the bound in Proposition III.2.

Proof of Proposition F.2. Given the tensor product input state ρ and for any $O \in \mathcal{O}_V$, we consider the variance

$$\text{Var}_{\rho_{\text{out}}}(O) = \text{tr}[\mathcal{N}_V(\rho) (O - \text{tr}[\mathcal{N}_V(\rho) O] \, \mathbb{I})^2]$$
$$= \text{tr}[(\rho \otimes \sigma_A) \mathcal{U}_{VA}^\dagger (O - \text{tr}[(\rho \otimes \sigma_A) \mathcal{U}_{VA}^\dagger(O)] \mathbb{I})^2]$$
$$= \text{Var}_{\rho \otimes \sigma_A}[\mathcal{U}_{VA}^\dagger(O)].$$

Next, we consider the Lipschitz constant $\| \cdot \|_L^{(VA)}$ corresponding to decomposition (F2). In particular, system $\mathcal{H}_V \otimes \mathcal{H}_A$ is constituted of at most $|V|(1 + L/2)$ particles. Since state $\rho \otimes \sigma_A$ is a tensor product state with respect to the above decomposition, we have the following $(2, \infty)$-Poincaré inequality from Lemma F.1:

$$\text{Var}_{\rho_{\text{out}}}(O) \leq \sum_{\omega \in VA} (\partial_\omega \mathcal{U}_{VA}^\dagger(O))^2 = \sum_{v \in V} (\partial_v \mathcal{U}_{VA}^\dagger(O))^2$$
$$+ \sum_{a \in A} (\partial_a \mathcal{U}_{VA}^\dagger(O))^2.$$

For the first sum, we obtain, using Proposition F.1,

$$\sum_{v \in V} (\partial_v \mathcal{U}_{VA}^\dagger(O))^2 \leq 4 \sum_{v \in V} \left(\sum_{\omega \in I_v^{\mathcal{U}_{VA}}} \partial_\omega O \right)^2$$
$$= 4 \sum_{v \in V} \left(\sum_{\omega \in I_v^{\mathcal{U}_{VA}} \backslash A} \partial_\omega O \right)^2$$
$$\leq 4 |V| \max_{v \in V} |I_v^{\mathcal{U}_{VA}} \backslash A|^2 \|O\|_L^2$$
$$= 4 |V| \max_{v \in V} |I_v^{\mathcal{N}_V}|^2 \|O\|_L^2$$
$$= 4 |V| I_{\mathcal{N}_V}^2 \|O\|_L^2.$$

The second sum on the other hand can be controlled as follows. First, for any layer ancilla a, denote by ℓ_a the layer

at which a is brought, and decompose the dilation $\mathcal{U}_{VA}$ as

$$\mathcal{U}_{VA} = \mathcal{U}_{VA}^{[\ell_a, L]} \circ \mathcal{U}_{VA}^{[1, \ell_a - 1]},$$

where the first subcircuit $\mathcal{U}_{VA}^{[1, \ell_a - 1]}$ corresponds to the first $\ell_a - 1$ layers, and the second subcircuit $\mathcal{U}_{VA}^{[\ell_a, L]}$ corresponds to the other layers. Then,

$$\partial_a \mathcal{U}_{VA}^\dagger(O) = 2 \min_{O_{a^c}} \|(\mathcal{U}_{VA}^{[1, \ell_a - 1]})^\dagger \circ (\mathcal{U}_{VA}^{[\ell_a, L]})^\dagger(O)$$
$$- O_{a^c} \otimes \mathbb{I}_a\|_\infty \leq \partial_a (\mathcal{U}_{VA}^{[\ell_a, L]})^\dagger(O),$$

where the inequality arises by choosing $O_{a^c} = (\mathcal{U}_{VA}^{[1, \ell_a - 1]})^\dagger(\widetilde{O}_{a^c})$ with $\widetilde{O}_{a^c}$ the optimizer of $\partial_a(\mathcal{U}_{VA}^{[\ell_a, L]})^\dagger(O)$. Therefore,

$$\sum_{a \in A} (\partial_a \mathcal{U}_{VA}^\dagger(O))^2 \leq \sum_{a \in A} ((\mathcal{U}_{VA}^{[\ell_a, L]})^\dagger(O))^2$$
$$\leq 4 \sum_{a \in A} \left(\sum_{\omega \in I_a^{\mathcal{U}_{VA}^{[\ell_a, L]}} \backslash A} \partial_\omega O \right)^2$$
$$\leq 4 \|O\|_L^2 \sum_{a \in A} |I_a^{\mathcal{U}_{VA}^{[\ell_a, L]}} \backslash A|^2$$
$$\leq 4 \|O\|_L^2 \max_\ell |E_\ell| \sum_{\ell=1}^L \max_{e \in E_\ell} I(e, L - \ell)^2,$$

where the second inequality again follows from Proposition F.1. ∎

APPENDIX G: CONTROLLING THE LIPSCHITZ CONSTANT FOR HAMILTONIAN DYNAMICS

In this appendix, we prove Eq. (17) (see Ref. [45, Proposition B.2.] for a similar statement). We first recall the following equivalent formulation of the Wasserstein distance due to Palma *et al.* [29]:

$$W_1(\rho, \sigma) = \frac{1}{2} \min \left\{ \sum_{i=1}^n \|X^{(i)}\|_1 : \rho - \sigma \right.$$
$$\left. = \sum_{i=1}^n X^{(i)}, X^{(i)} \in \mathcal{O}_V^T, \text{tr}_i[X^{(i)}] = 0 \right\}. \quad \text{(G1)}$$

Here we recall that $\mathcal{O}_V^T$ denotes the set of self-adjoint, traceless observables.

Proposition G.1. *Assume that the continuous-time evolution $\{\mathcal{U}_V(t)\}_{t \geq 0}$ defined on the graph $G = (V, E)$ with $|V| = n$ satisfies the bound in Theorem III.2. Then, for any $H \in \mathcal{O}_V$,*

$$\|\mathcal{U}_V(t)^\dagger(H)\|_L \leq \left(2(i_0 - 1) + \frac{4M}{2D - 1} \sum_{i=i_0}^{n} d(i)^{\delta-1} e^{vt - d(i)}\right)\|H\|_L, \tag{G2}$$

where $\mathrm{dist}(\{1\}, \{i, \ldots, n\}) \equiv d(i)$*, and* i_0 *stands for the first vertex such that* $d(i_0) \geq 2\delta - 1$.

Proof. From Ref. [29], the Wasserstein distance W_1 arises from a norm $\|\cdot\|_{W_1}$, i.e., $W_1(\rho, \sigma) = \|\rho - \sigma\|_{W_1}$. Moreover, the norm $\|\cdot\|_{W_1}$ is uniquely determined by its unit ball, which in turn is the convex hull of the set of differences between couples of neighboring quantum states:

$$\mathcal{N}_n = \bigcup_{i \in V} \mathcal{N}_n^{(i)}, \qquad \mathcal{N}_n^{(i)} = \{\rho - \sigma : \rho, \sigma \in \mathcal{S}_V, \, \mathrm{tr}_i(\rho) = \mathrm{tr}_i(\sigma)\}.$$

Now, by convexity, the contraction coefficient for this norm is equal to

$$\|\mathcal{U}_V(t)\|_{W_1 \to W_1} = \max\{\|\mathcal{U}_V(t)(X)\|_{W_1} : X \in \mathcal{O}_V^T, \|X\|_{W_1} \leq 1\} = \max_{X \in \mathcal{N}_n} \|\mathcal{U}_V(t)(X)\|_{W_1}.$$

Now let $X \in \mathcal{N}_n$. By expression (G1), and choosing without loss of generality an ordering of the vertices such that $\mathrm{tr}_1(X) = 0$, we have

$$\begin{aligned}
\|\mathcal{U}_V(t)(X)\|_{W_1} &\leq \frac{1}{2} \sum_{i=1}^{n} \left\| \frac{I}{d^{i-1}} \otimes \mathrm{tr}_{1\cdots i-1} \circ \mathcal{U}_V(t)(X) - \frac{I}{d^i} \otimes \mathrm{tr}_{1\cdots i} \circ \mathcal{U}_V(t)(X) \right\|_1 \\
&= \frac{1}{2} \sum_{i=1}^{n} \left\| \int d\mu(U_i) \, \mathrm{tr}_{1\cdots i-1} \circ (\mathcal{U}_V(t)(X) - U_i \mathcal{U}_V(t)(X) U_i^\dagger) \right\|_1 \\
&\leq \frac{1}{2} \sum_{i=1}^{n} \int d\mu(U_i) \|[U_i, \mathrm{tr}_{1\cdots i-1} \circ \mathcal{U}_V(t)(X)]\|_1 \\
&\leq \sum_{i=1}^{n} \| \mathrm{tr}_{1\cdots i-1} \circ \mathcal{U}_V(t)(X)\|_1 \\
&\overset{(1)}{=} \sum_{i=1}^{n} \| \mathrm{tr}_{1\cdots i-1} \circ (\mathcal{U}_V(t) - \mathcal{U}_{\{i-k,\ldots,n\}}(t))(X)\|_1, \tag{G3}
\end{aligned}$$

where μ denotes the Haar measure on one qudit, and where equality (1) follows from the fact that $\mathrm{tr}_1(X) = 0$, with $\mathcal{U}_{\{i-k,\ldots,n\}}(t)$ defined as in Theorem III.2 with $k < i - 1$. Next, by the variational formulation of the trace distance and Theorem III.2, we have, for $i \geq i_0$,

$$\begin{aligned}
\| \mathrm{tr}_{1\cdots i-1} \circ (\mathcal{U}_V(t) - \mathcal{U}_{\{i-k,\ldots,n\}}(t))(X)\|_1 &= \max_{\|O_{i\cdots n}\|_\infty \leq 1} |\mathrm{tr}[X(\mathcal{U}_V(t)^\dagger - \mathcal{U}_{\{i-k,\ldots,n\}}(t)^\dagger)(O_{i\cdots n})]| \\
&\leq \max_{\|O_{i\cdots n}\|_\infty \leq 1} \|(\mathcal{U}_V(t)^\dagger - \mathcal{U}_{\{i-k,\ldots,n\}}(t)^\dagger)(O_{i\cdots n})\|_\infty \|X\|_1 \\
&\leq \frac{2M}{2D - 1} d_{i,k}^{\delta-1} e^{vt - d_{i,k}} \|X\|_1 \\
&\overset{(2)}{\leq} \frac{4M}{2D - 1} d_{i,k}^{\delta-1} e^{vt - d_{i,k}} \|X\|_{W_1},
\end{aligned}$$

where $d_{i,k} := \mathrm{dist}(\{i \cdots n\}, \{1 \cdots i - k - 1\})$ and inequality (2) follows from Ref. [29, Proposition 6]. By picking $k = i - 2$ and inserting this estimate into Eq. (G3) for $i \geq i_0$ and the trivial estimate $\| \mathrm{tr}_{1\cdots i-1} \circ (\mathcal{U}_V(t) - \mathcal{U}_{\{i-k,\ldots,n\}}(t))(X)\|_1 \leq 2\|X\|_1$ for $i \leq i_0 - 1$, we obtain Eq. (G2) by duality. ∎

APPENDIX H: PROPERTIES OF HIGH-ENERGY STATES OF THE MAX-CUT HAMILTONIAN

Proposition H.1. *Under the same hypotheses as Theorem VI.1, let x_{opt} achieve the maximum cut of G, and let X be the random outcome obtained measuring ρ in the computational basis. Then,*

$$\mathbb{P}\left(d_H(X, x_{\text{opt}}) \leq \frac{n}{3}\right) = \mathbb{P}\left(d_H(X, \bar{x}_{\text{opt}}) \leq \frac{n}{3}\right) \geq \frac{1}{4}, \tag{H1}$$

where $\bar{x}$ denotes the bitwise negation of x.

Proof. The proof follows the same lines as the proof of Theorem 1 of Ref. [16]. Since G is regular and bipartite, we have $2|E| = Dn$ and $C_{\text{max}} = |E|$, and Eq. (40) becomes

$$\text{tr}[\rho H] \geq |E| - \frac{hn}{6}. \tag{H2}$$

We have

$$\text{tr}[\rho H] = \mathbb{E}C(X), \tag{H3}$$

and Markov's inequality implies that

$$\mathbb{P}\left(C(X) \geq |E| - \frac{h\,n}{3}\right) \geq \frac{1}{2}. \tag{H4}$$

Since G is bipartite, we have, for any $x \in \{0,1\}^n$,

$$C(x) + C(x_{\text{opt}} \oplus x) = |E|, \tag{H5}$$

where "$\oplus$" denotes the sum modulo 2. Equations (H4) and (H5) imply that

$$\mathbb{P}\left(C(x_{\text{opt}} \oplus X) \leq \frac{hn}{3}\right) \geq \frac{1}{2}. \tag{H6}$$

Hypotheses (38) and (H6) imply that

$$\mathbb{P}\left(d_H(X, x_{\text{opt}}) \leq \frac{n}{3}\right) + \mathbb{P}\left(d_H(X, \bar{x}_{\text{opt}}) \leq \frac{n}{3}\right) \geq \frac{1}{2}, \tag{H7}$$

where we have used the fact that, for any x, $y \in \{0,1\}^n$,

$$|x \oplus y| = d_H(x, y), \qquad d_H(x, y) + d_H(\bar{x}, y) = n. \tag{H8}$$

Since ρ commutes with $\sigma_x^{\otimes n}$, the probability distribution of X is invariant with respect to the negation of all the bits; therefore,

$$\mathbb{P}\left(d_H(X, x_{\text{opt}}) \leq \frac{n}{3}\right) = \mathbb{P}\left(d_H(X, \bar{x}_{\text{opt}}) \leq \frac{n}{3}\right). \tag{H9}$$

The claim follows. ∎

APPENDIX I: CONCENTRATION FOR ERROR MITIGATION

In this section we prove Theorem V.1, which we restate for the reader's convenience.

Theorem I.1. *For an error-mitigation observable X,*

$$X := \sum_{s \in \mathcal{S}} f(s) \text{tr}_A(I_S \otimes |0\rangle \langle 0|^{\otimes k} M_s), \tag{I1}$$

assume that, for a given state σ,

$$\mathbb{P}_{\sigma^{\otimes m}}(\|s - \mathbb{E}_{\sigma^{\otimes m}}(s)\|_{\ell_\infty} \geq rn) \leq K(m) \exp\left(-\frac{cr^2 n}{\ell_0^2}\right)$$

holds for some function $K(m)$. Furthermore, assume that, for $r, \epsilon > 0$, given that, for all $1 \leq i \leq m$, $D_2(\mathcal{E}_i(\rho)\|\sigma) = c(r - \epsilon)n/ml_0$ and f is L_f Lipschitz with respect to the ℓ_∞ norm, then, for $\rho_{\text{out}} = \bigotimes_{i=1}^m \mathcal{E}_i(\rho)$, we have

$$\mathbb{P}_{\rho_{\text{out}}}(|X - f(\mathbb{E}_{\sigma^{\otimes m}}(s))\mathbb{I}| > rL_f n) \leq \exp\left(-\frac{c\epsilon n}{\ell_0^2}\right). \tag{I2}$$

Proof. It follows from the additivity of the Rényi divergence and our assumption on $D_2(\mathcal{E}_i(\rho)\|\sigma)$ that

$$D_2\left(\bigotimes_{i=1}^m \mathcal{E}_i(\rho)\,\bigg\|\,\sigma^{\otimes m}\right) = \sum_i D_2(\mathcal{E}_i(\rho)\|\sigma) \leq \frac{c(r^2 - \epsilon)}{l_0^2}. \tag{I3}$$

From Lemma B.1 and the concentration inequality in Eq. (I2) we obtain

$$\mathbb{P}_{\rho_{\text{out}}}(\|s - \mathbb{E}_{\sigma^{\otimes m}}(s)\|_{\ell_\infty} \geq rn) \leq K(m) \exp\left(-\frac{c\epsilon}{\ell_0^2}\right).$$

By our assumption on function f being Lipschitz, we have

$$\mathbb{P}_{\rho_{\text{out}}}(|X - f(\mathbb{E}_{\sigma^{\otimes m}}(s))\mathbb{I}| \leq rL_f n)$$
$$\geq \mathbb{P}_{\rho_{out}}(\|s - \mathbb{E}_{\sigma^{\otimes m}}(s)\|_{\ell_\infty} \leq rn),$$

and the claim follows. ∎

[1] F. Arute *et al.*, Quantum supremacy using a programmable superconducting processor, Nature **574**, 505 (2019).

[2] S. Ebadi, T. T. Wang, H. Levine, A. Keesling, G. Semeghini, A. Omran, D. Bluvstein, R. Samajdar, H. Pichler, W. W. Ho, S. Choi, S. Sachdev, M. Greiner, V. Vuletić, and M. D. Lukin, Quantum phases of matter on a 256-atom programmable quantum simulator, Nature **595**, 227 (2021).

[3] P. Scholl, M. Schuler, H. J. Williams, A. A. Eberharter, D. Barredo, K. N. Schymik, V. Lienhard, L. P. Henry, T. C. Lang, T. Lahaye, A. M. Lauchli, and A. Browaeys, Quantum simulation of 2D antiferromagnets with hundreds of rydberg atoms, Nature **595,** 233 (2021).

[4] H. S. Zhong *et al.*, Quantum computational advantage using photons, Science **370,** 1460 (2020).

[5] E. T. Campbell, B. M. Terhal, and C. Vuillot, Roads towards fault-tolerant universal quantum computation, Nature **549,** 172 (2017).

[6] J. Roffe, Quantum error correction: An introductory guide, Contemp. Phys. **60,** 226 (2019).

[7] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2,** 79 (2018).

[8] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum algorithms, Rev. Mod. Phys. **94,** 015004 (2022).

[9] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3,** 625 (2021).

[10] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5,** 4213 (2014).

[11] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm. arXiv preprint arXiv:1411.4028 (Nov. 2014).

[12] E. R. Anschuetz, in *International Conference on Learning Representations*, (2022).

[13] L. Bittel and M. Kliesch, Training Variational Quantum Algorithms is NP-Hard, Phys. Rev. Lett. **127,** 120502 (2021).

[14] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9,** 1 (2018).

[15] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, Nat. Commun. **12,** 6961 (2021).

[16] S. Bravyi, A. Kliesch, R. Koenig, and E. Tang, Obstacles to Variational Quantum Optimization from Symmetry Protection, Phys. Rev. Lett. **125,** 260505 (2020).

[17] C.-N. Chou, P. J. Love, J. S. Sandhu, and J. Shi, Limitations of local quantum algorithms on random Max-k-XOR and beyond, arXiv preprint arXiv:2108.06049 (2021).

[18] E. Farhi, D. Gamarnik, and S. Gutmann, The quantum approximate optimization algorithm needs to see the whole graph: A typical case. arXiv:2004.09002 (2020).

[19] A. H. Moosavian, S. S. Kahani, and S. Beigi, Limits of short-time quantum annealing, arXiv:2104.12808 [quant-ph] (2021).

[20] D. S. França and R. Garcia-Patrón, Limitations of optimization algorithms on noisy quantum devices, Nat. Phys. **17,** 1221 (2021).

[21] G. González-García, R. Trivedi, and J. I. Cirac, Error propagation in NISQ devices for solving classical optimization problems, arXiv preprint arXiv:2203.15632 (2022).

[22] E. A. Carlen and J. Maas, An analog of the 2-Wasserstein metric in non-commutative probability under which the fermionic Fokker–Planck equation is gradient flow for the entropy, Commun. Math. Phys. **331,** 887 (2014).

[23] E. A. Carlen and J. Maas, Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance, J. Funct. Anal. **273,** 1810 (2017).

[24] E. A. Carlen and J. Maas, Non-commutative calculus, optimal transport and functional inequalities in dissipative quantum systems, J. Stat. Phys. **178,** 319 (2019).

[25] N. Datta and C. Rouzé, Relating relative entropy, optimal transport and Fisher information: A quantum HWI inequality, Annales Henri Poincaré **21,** 2115 (2020).

[26] G. De Palma and C. Rouzé, Quantum concentration inequalities, arXiv preprint arXiv:2106.15819 (2021).

[27] L. Gao, M. Junge, and N. LaRacuente, Fisher information and logarithmic Sobolev inequality for matrix-valued functions, Annales Henri Poincaré **21,** 3409 (2020).

[28] L. Gao and C. Rouzé, Ricci curvature of quantum channels on non-commutative transportation metric spaces, arXiv preprint arXiv:2108.10609 (2021).

[29] G. D. Palma, M. Marvian, D. Trevisan, and S. Lloyd, The quantum Wasserstein distance of order 1, IEEE Trans. Inf. Theory **67,** 6627 (2021).

[30] G. D. Palma and D. Trevisan, Quantum optimal transport with quantum channels, Annales Henri Poincaré **22,** 3199 (2021).

[31] C. Rouzé and N. Datta, Concentration of quantum states from quantum functional and transportation cost inequalities, J. Math. Phys. **60,** 012202 (2019).

[32] L. Eldar and A. W. Harrow, in *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, Berkeley, CA, 2017), p. 427.

[33] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, Virtual Distillation for Quantum Error Mitigation, Phys. Rev. X **11,** 041036 (2021).

[34] B. Koczor, Exponential Error Suppression for Near-Term Quantum Devices, Phys. Rev. X **11,** 031057 (2021).

[35] R. Takagi, S. Endo, S. Minagawa, and M. Gu, Fundamental limits of quantum error mitigation, arXiv preprint arXiv:2109.04457 (Sept. 2021).

[36] S. Wang, P. Czarnik, A. Arrasmith, M. Cerezo, L. Cincio, and P. J. Coles, Can error mitigation improve trainability of noisy variational quantum algorithms? arXi preprint arXiv:2109.01051 (Sept. 2021).

[37] E. Milman, On the role of convexity in isoperimetry, spectral-gap and concentration, Invent. Math. **177,** 01 (2008).

[38] B. Jourdain, Equivalence of the Poincaré inequality with a transport-chi-square inequality in dimension one, Electron. Commun. Probab. **17,** 1 (2012).

[39] M. Ledoux, Remarks on some transportation cost inequalities. *Unpublished notes available on the author's website* https://perso.math.univ-toulouse.fr/ledoux/files/2019/03/transport.pdf, (2018).

[40] Y. Liu, The Poincaré inequality and quadratic transportation-variance inequalities, Electron. J. Probab. **25**, 1 (2020).

[41] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, On quantum Rényi entropies: A new generalization and some properties, J. Math. Phys. **54**, 122203 (2013).

[42] M. M. Wilde, A. Winter, and D. Yang, Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy, Commun. Math. Phys. **331**, 593 (2014).

[43] S. Bobkov and F. Götze, Exponential integrability and transportation cost related to logarithmic Sobolev inequalities, J. Funct. Anal. **163**, 1 (1999).

[44] K. Marton, A simple proof of the blowing-up lemma (corresp.), IEEE Trans. Inf. Theory **32**, 445 (1986).

[45] C. Rouzé and D. S. França, Learning quantum many-body systems from a few copies, arXiv preprint arXiv:2107.03333 (2021).

[46] F. G. S. L. Brandao and M. Cramer, Equivalence of statistical mechanical ensembles for non-critical quantum systems, (2015).

[47] T. Kuwahara and K. Saito, Eigenstate Thermalization from the Clustering Property of Correlation, Phys. Rev. Lett. **124**, 200604 (2020).

[48] H. Tasaki, On the local equivalence between the canonical and the microcanonical ensembles for quantum spin systems, J. Stat. Phys. **172**, 905 (2018).

[49] E. H. Lieb and D. W. Robinson, in *Statistical mechanics* (Springer, Berlin, Heidelberg, 1972), p. 425.

[50] K. Temme, M. J. Kastoryano, M. B. Ruskai, M. M. Wolf, and F. Verstraete, The $\chi 2$-divergence and mixing times of quantum Markov processes, J. Math. Phys. **51**, 122201 (2010).

[51] B. Nachtergaele and R. Sims, Lieb-Robinson bounds and the exponential clustering theorem, Commun. Math. Phys. **265**, 119 (2006).

[52] T. Barthel and M. Kliesch, Quasilocality and Efficient Simulation of Markovian Quantum Dynamics, Phys. Rev. Lett. **108**, 230504 (2012).

[53] M. Kliesch, C. Gogolin, and J. Eisert, in *Many-Electron Approaches in Physics, Chemistry and Mathematics* (Springer, Cham, 2014), p. 301.

[54] S. Beigi, N. Datta, and C. Rouzé, Quantum reverse hypercontractivity: Its tensorization and application to strong converses, Commun. Math. Phys. **376**, 753 (2020).

[55] Ángela Capel, C. Rouzé and D. S. França, The modified logarithmic Sobolev inequality for quantum spin systems: Classical and commuting nearest neighbour interactions, arXiv preprint arXiv:2009.11817 (Sept. 2020).

[56] T. Kuwahara and K. Saito, Gaussian concentration bound and ensemble equivalence in generic quantum many-body systems including long-range interactions, Ann. Phys. (N. Y) **421**, 168278 (2020).

[57] A. Anshu, Concentration bounds for quantum states with finite correlation length on quantum spin lattice systems, New J. Phys. **18**, 083011 (2016).

[58] M. Berta, D. Sutter, and M. Walter, Quantum Brascamp-Lieb dualities, arXiv:1909.02383 (2019).

[59] R. Carbone and A. Martinelli, Logarithmic Sobolev inequalities in non-commutative algebras, Infinite Dimensional Anal., Quantum Probability Relat. Top. **18**, 1550011 (2015).

[60] L. Gao and C. Rouzé, Complete entropic inequalities for quantum Markov chains, (2021).

[61] C. Hirche, C. Rouzé, and D. S. França, On contraction coefficients, partial orders and approximation of capacities for quantum channels, arXiv preprint arXiv:2011.05949 (Nov. 2020).

[62] M. J. Kastoryano and K. Temme, Quantum logarithmic sobolev inequalities and rapid mixing, J. Math. Phys. **54**, 052202 (2013).

[63] A. Müller-Hermes and D. S. Franca, Sandwiched Rényi convergence for quantum evolutions, Quantum **2**, 55 (2018).

[64] R. Olkiewicz and B. Zegarlinski, Hypercontractivity in noncommutative LpSpaces, J. Funct. Anal. **161**, 246 (1999).

[65] X. Bao, N. V. Sahinidis, and M. Tawarmalani, Semidefinite relaxations for quadratically constrained quadratic programming: A review and comparisons, Math. Program. **129**, 129 (2011).

[66] S. Endo, S. C. Benjamin, and Y. Li, Practical Quantum Error Mitigation for Near-Future Applications, Phys. Rev. X **8**, 031027 (2018).

[67] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, Hybrid quantum-classical algorithms and quantum error mitigation, J. Phys. Soc. Jpn **90**, 032001 (2021).

[68] S. McArdle, X. Yuan, and S. Benjamin, Error-Mitigated Digital Quantum Simulation, Phys. Rev. Lett. **122**, 180501 (2019).

[69] K. Temme, S. Bravyi, and J. M. Gambetta, Error Mitigation for Short-Depth Quantum Circuits, Phys. Rev. Lett. **119**, 180509 (2017).

[70] M. X. Goemans and D. P. Williamson, Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming, J. ACM (JACM) **42**, 1115 (1995).

[71] A. Marcus, D. A. Spielman, and N. Srivastava, in *2013 IEEE 54th Annual Symposium on Foundations of computer science* (IEEE, Berkeley, CA, 2013), p. 529.

[72] A. W. Marcus, D. A. Spielman, and N. Srivastava, Interlacing families IV: Bipartite Ramanujan graphs of all sizes, SIAM J. Comput. **47**, 2488 (2018).

[73] M. Morgenstern, Existence and explicit constructions of q + 1 regular Ramanujan graphs for every prime power q, J. Comb. Theory, Ser. B **62**, 44 (1994).

[74] J. Friedman, *A proof of Alon's second eigenvalue conjecture and related problems* (American Mathematical Soc., New York, NY, 2008).

[75] S. Jansen, M.-B. Ruskai, and R. Seiler, Bounds for the adiabatic approximation with applications to quantum computation, J. Math. Phys. **48**, 102111 (2007).

[76] J. Basso, E. Farhi, K. Marwaha, B. Villalonga, and L. Zhou, The quantum approximate optimization algorithm at high depth for maxcut on large-girth regular graphs and the Sherrington-Kirkpatrick model. *In Proceedings of the 17th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC '22), 7:1–7:21, (2022)*, (Oct. 2021).

[77] A. Anshu and T. Metger, Concentration bounds for quantum states and limitations on the QAOA from polynomial approximations. arXiv preprint arXiv:2209.02715, (2022).

[78] M. Christandl and A. Müller-Hermes, Relative entropy bounds on quantum, private and repeater capacities, Commun. Math. Phys. **353**, 821 (2017).

[79] K. Temme, F. Pastawski, and M. J. Kastoryano, Hypercontractivity of quasi-free quantum semigroups, J. Phys. A: Math. Theor. **47**, 405303 (2014).

[80] J. Basso, E. Farhi, K. Marwaha, B. Villalonga, and L. Zhou, The quantum approximate optimization algorithm at high depth for Maxcut on large-girth regular graphs and the Sherrington-Kirkpatrick model, (2021).

[81] E. Farhi, J. Goldstone, S. Gutmann, and L. Zhou, The quantum approximate optimization algorithm and the Sherrington-Kirkpatrick model at infinite size, arXiv:1910.08187 (2019).

[82] A. Dembo, A. Montanari, and S. Sen, Extremal cuts of sparse random graphs. *Annals of Probability, 2017, Vol 45, No. 2, 1190-1217*, (Mar. 2015).

[83] G. Parisi, Toward a mean field theory for spin glasses, Phys. Lett. A **73**, 203 (1979).

[84] J. K. Thompson, O. Parekh, and K. Marwaha, An explicit vector algorithm for high-girth maxcut. *In Proceedings of the Symposium on Simplicity in Algorithms (SOSA 2022); pp. 238-246*, (2021).